University of Redlands

**Economic Analysis Tool for
The California Budget**

A Major Individual Project submitted in partial satisfaction of the requirements
for the degree of Master of Science in Geographic Information Systems

by

April Marie Villagomez-Roe

Douglas Flewelling Ph.D., Committee Chair

Mark Kumler, Ph.D.

January 2012

Economic Analysis Tool for
The California Budget

Copyright © 2012

by

April Marie Villagomez-Roe

The report of April Marie Villagomez-Roe is approved.

Mark Kumler, Ph.D.

Douglas Flewelling, Ph.D., Committee Chair

January 2012

Acknowledgements

First and foremost, I would like to thank God for giving me the opportunity to pursue my dreams and aspirations. None of this would have been possible and I am grateful to Him.

To my husband, Jarod, no words can express how appreciative I am since you were able to stand by and support me emotionally and financially. When no one believed in me, you did. When times were rough with life, you were there for me. You were not the least bit selfish and were able to sacrifice not having your wife for a full year. I hope the time spent away from each other will prove to benefit the both of us and our future family. I love you so much, and I hope one day I am able to return the same love and support to you.

To my parents, Elias and Cecilia, and Sister Jennifer, thank you for letting me move back in temporarily while I pursued my degree. I enjoyed the time we spent together, the yummy food, and most importantly, the love and support that you supplied all year long. You have helped Jarod and I tremendously, and we cannot say thank you enough. I could not ask for a more supportive and loving family, and hope to repay you in some way in the near future.

To Dr. Douglas Flewelling, thank you for agreeing to take me under your knowledgeable wing and give me the confidence I needed to believe that I could succeed in completing this program. Although I was nervous at first to have the Director of the MS GIS program as my advisor, I hope I was able to surpass your expectations and hope you were still confident in your decision of having me as an advisee. I have enjoyed the weekly meetings we had because even if we had simple conversations, I felt that I was constantly learning from you. Not very many Professors do that for students these days, but I am extremely appreciative of the numerous opportunities you have given me to learn more and think “outside the box.” Last but not least, thank you for the scholarships you bestowed upon me. I cannot express how much it meant to Jarod and I that you were able to help us out financially. I plan to return the favor for the next student who is in financial need in the near future.

To Debra Riley, I appreciate you very much. Since our first conversation back in 2009, I was excited to learn what the program had offered and you were there answering the hundreds of questions I had. I am certainly glad that you convinced me to not start a family while pursuing a Master’s degree. I would have been an absolute wreck and my studies would have suffered. In the end, you were supportive, in your own way, and this is why “Den Mother” suits you as a perfect nickname!

To Dr. Mark Kumler, thank you for having a great sense of humor as Andre and I had our way with your personal property, and allowing me to photograph you while you were sleeping in every inopportune place. You have been a great listener to the majority of Cohort 19 when times were rough. I doubt you will remember what was said, but when we needed you, you were there for us. If somehow my career in GIS does not work out in the near future, I will pursue a professional career in shuffleboard and take you on as my coach. Thanks for introducing us to Hanger 24; it was a definite staple in our stressful graduate student lives! P.S.: sorry my MIP defense was so boring that you felt the need to “rest your eyes!”

Last and certainly not least, huge thanks to my two greatest friends, LeAndre Roberson and Marc Guidry. We had so many great memories of being “locked up” in Lewis Hall 103. Watching basketball, baseball, football games on the big screen with pizza and “drinks” are one of the many things I will remember from our time together. I am glad we were such great support systems for each other, and hope we can continue to be for many years to come. Andre, you and I have had a great friendship since day one. I will always be grateful to you because you helped support me in so many ways, especially since I was a broke graduate student. We shared many worries, strife, goals and things that made us happy. Your support has helped me in so many ways this past year, I hope to return the favor one day, but until then all I can say is thank you!

Abstract

Economic Analysis Tool for the California Budget

by
April Marie Villagomez-Roe

California's deficit has been a political issue for many years. High levels of state spending, and a high unemployment rate due to the current recession, are contributing factors to this problem. Since California voters disapprove of any tax increase to balance the budget, a solution to the deficit needs an alternative approach. To look at the possible spatial effects of any changes to the yearly budget, a geographic information system (GIS) was used to build a model to help in developing a plan to spread the budget cuts more evenly across California. Analyzing these effects could help policy makers in developing the spatial distribution plan of the budget cuts since the State government does not want to increase economic turmoil. As a demonstration, the budget analyst tool is used to calculate potential effects of income and sales tax increases using 2009-2010 economic data. The user can make assumptions as to how to spread tax increases evenly throughout the state by running the model with different tax rates until the desired effects are achieved. The results are used to view the effects of the tax increases and understand the spatial pattern associated with the increases.

Table of Contents

Chapter 1 – Introduction	1
1.1 Client.....	1
1.2 Problem Statement.....	1
1.3 The Implemented System	1
1.3.1 Goals and Objectives	2
1.3.2 Scope.....	2
1.3.3 Methods.....	3
1.4 Audience	4
1.5 Overview of the Rest of this Report	4
Chapter 2 – Background and Literature Review	5
2.1 GIS and Decision Making in the Government Sector	5
2.2 GIS and the Geography of Politics	7
2.3 The Effects of Redistricting Political Boundaries	8
2.4 Analyzing Demographic and Economic Conditions.....	9
2.5 Summary	11
Chapter 3 – Systems Analysis and Design.....	13
3.1 Problem Statement.....	13
3.2 Requirements Analysis	13
3.2.1 Functional Requirements	15
3.2.2 Non-Functional Requirements	16
3.3 System Design	17
3.3.1 Process Component.....	17
3.3.2 Hardware Component	18
3.3.3 Software Component	18
3.3.4 Data Component	18
3.4 Project Plan	19
3.5 Summary	20
Chapter 4 – Database Design.....	21
4.1 Conceptual Data Model	21
4.2 Logical Data Model	22
4.3 Data Sources	24
4.4 Data Scrubbing and Loading	24
4.5 Summary	26
Chapter 5 – Implementation.....	27
5.1 State Budget Data Exploration.....	27
5.2 Data Conversion.....	27
5.3 Data Refinement	28
5.4 Data Model.....	31
5.5 Summary	34
Chapter 6 – Results and Analysis.....	37
6.1 Budget Analyst Tool County Output Results	37

6.2	Budget Analyst Tool ZIP Code Output Results.....	41
6.3	Summary	45
Chapter 7 –	Summary and Future Work.....	47
7.1	Summary of Objectives Achieved	47
7.2	Future Work	47
7.2.1	Expansion of Budget Analyst Tool.....	47
7.2.2	Budget Expenditure Tool.....	47
7.2.3	Customized Budget Analyst Tool Website.....	48
7.2.4	Addition of Sales Tax Industry Precision	48
7.3	Summary	48
Works Cited	49	
Appendix A.	County Budget Analyst Tool Python Script	50
Appendix B.	ZIP Code Budget Analyst Tool Python Script.....	52

Table of Figures

Figure 1.1: Example Map Displaying Sales Tax Increase Normalized By Population.	3
Figure 2.1: Potential Workflow of Decision Making in the Government Sector.	5
Figure 2.2: Range of GIS Uses for Decision Making (Adapted from Berry, 1994).	6
Figure 2.3: Website of County Comparison Maps based on Education Funding and Population Under 18 (United States Government, 2011).	8
Figure 2.4: Gerrymandered Fourth Congressional District of Chicago, Illinois (Office of Senate Counsel & Research for the Minnesota Senate, 2002).	9
Figure 2.5: Subjects included in the American Community Survey (United States Census Bureau, 2008).	10
Figure 3.1: Project Logic Diagram (Tomlinson, 2007).	14
Figure 3.2: System Design Components.	17
Figure 3.3: Original and Revised Project Schedules.	19
Figure 4.1: California State Budget Conceptual Model.	21
Figure 4.2: Data Retrieval Method.	25
Figure 5.1: Franchise Tax Board ZIP Code Income Tax Statistics as a PDF with Data Converted to Excel Format Using Able2Extract Professional Software.	28
Figure 5.2: California State Board of Equalization Data as Downloaded with Data Reformatted for Import into ArcGIS.	29
Figure 5.3: U.S. Bureau of Economic Analysis Data as Downloaded with Per Capita Personal Income Data Reformatted for Import into ArcGIS.	30
Figure 5.4: General Flow of County and ZIP Code Data Models.	31
Figure 5.5: Pop-Up Window Display for User Sales Tax Percentage Input into the Model.	31
Figure 5.6: Budget Analysis Equation Provided by the Client.	32
Figure 5.7: Adjusted Equation Used in the Data Model.	32
Figure 5.8: Tool Output Accessibility from the Results Window (Esri, 2010).	33
Figure 5.9: Spatial Autocorrelation HyperText Markup Language (HTML) Report (Esri, 2011).	34
Figure 6.1: Budget Analyst Tool County Output Feature Class.	37
Figure 6.2: User Input Sales Tax Rate Revenue Results Shown in Map Template.	38
Figure 6.3: Derived Income Tax Rate Revenue Results Shown in Map Template.	39
Figure 6.4: Bivariate Map of Sales and Income Tax Per Capita Revenue Results.	40
Figure 6.5: Budget Analyst Tool ZIP Code Output Feature Class.	41
Figure 6.6: Venn Diagram of Approximate Data Matched Using Esri ZIP Code and Franchise Tax Board ZIP Code Datasets.	42
Figure 6.7: User Input Sales Tax Rate Revenue Results at ZIP Code Level.	43
Figure 6.8: Derived Income Tax Rate Revenue Results at ZIP Code Level.	44
Figure 6.9: A ZIP Code Area of San Francisco Shown with Business Locations.	45

List of Tables

Table 1. Non-Functional Requirements 16

Table 2. California State Budget Logical Data Model..... 23

List of Acronyms and Definitions

ACS	American Community Survey
BEA	Bureau of Economic Analysis
BOE	California State Board of Equalization
CAFR	Comprehensive Annual Financial Report
Esri	Environmental Systems Research Institute, Inc.
FTB	Franchise Tax Board
GIS	Geographic Information System
IPDs	Information Product Descriptions
IRS	Internal Revenue Service
PDF	Portable Document Format File

Chapter 1 – Introduction

The State of California is faced with creating a balanced budget every year. High levels of state spending, combined with high unemployment due to the current recession, have caused California to have the one of largest state deficits in the United States (Reuters, 2011) and this, in turn, effects the budget process. In 2011, Governor Jerry Brown proposed a budget dependent on deficit solutions such as reducing health and human services, education, and state programs and increasing revenue by realigning multiple funding sources or implementing tax increases. Since California voters disapproved of any tax increase, a solution to the deficit needed an alternative approach. To look at the regional effects of the yearly budget, the use of a geographic information system (GIS) as an analysis tool was constructed to help in developing a plan to spread the budget cuts evenly across California. Analyzing these effects would help policy makers in developing the spatial distribution plan of the budget cuts since the State government does not want to increase economic turmoil.

1.1 Client

Dr. Johannes Moenius, Associate Professor and Director of the Institute for Spatial Economic Analysis at the University of Redlands School of Business, was the client for this project. Dr. Moenius has extensive knowledge of international business and the global economy. Using spatial analysis tools, he has studied the Southern California housing market and the regional socioeconomic factors that fueled and popped the housing “bubble.” For this project, Dr. Moenius wanted to analyze the regional effects of the California state budget. Using the results, he would like to present to the state and local governments a plan on how to spread the financial burden of the state equitably.

1.2 Problem Statement

Policy makers need to visualize the regional impact of the tax increases and budget cuts. The client would like to help policy makers understand the implications of tax increases by showing areas of burden and the effects on various groups. In order for this to be achieved, a spatial budget analyst tool needed to be created with the use of available economic data. Since this is a spatial problem, identifying the primary burden per capita would assist in understanding which regions carry disproportional amounts of the deficit burden, and how it could be distributed more evenly throughout the state.

1.3 The Implemented System

The implemented system was centered on the output feature class of the budget analyst tool. The tool and data used for analysis were stored in the project geodatabase. When the tool was executed within the supplied map document, the user would enter a desired sales tax percentage and run the tool to produce an output feature class which was added to the map document. The output feature class contained the recommended income tax percentage rate, the revenue generated from both input and derived rates, along with additional economic data. The user was able to assign a classification symbology to the

output feature class and analyze the results. Once the analysis was complete, the map document contained a standardize map reporting template so the user could print individual maps of all fifty-eight counties. This ensured that maps printed contained a legend, title, and all other necessary map elements so that the user did not need to spend time on any cartographic changes.

1.3.1 Goals and Objectives

The goal of this project was to provide a budget analyst tool to help understand the effects of the state budget cuts and tax increases on regions and groups throughout California. The information provided by the analysis tool was used as an initial solution to decrease economic turmoil while California continues to work its way out of debt.

The major objectives of this project were the following:

1. Format and load all downloaded data in a geodatabase.
2. Incorporate the data in the modeling tool.
3. Display the data spatially in a map document.
4. Provide a standard reporting template.

Formatting the data from their original files to feature classes would expedite the loading of the data into a geodatabase. This would ensure that all the data used were stored in one location and maintained the same spatial projection. The data would be incorporated in the modeling tool to calculate the tax rate percentages that were input and derived. Once the calculated data have been output into a new feature class in geodatabase, the data would be displayed spatially and with ease for analysis. This would help the client view the effects of the state's tax increases, and the distribution of the financial burden. Lastly, a standard reporting template would be supplied for the user to print the maps without having to do any major cartography. This would allow the user to change any parameters in the modeling tool and produce maps for multiple regions in a single map document.

1.3.2 Scope

Multiple revenue sources and expenditures create the state budget proposal, which are aimed at decreasing the state's deficit. Since the project had data and time constraints, the scope focused on two revenue categories. The data were taken from the comprehensive annual financial reports downloaded from California's individual county websites and multiple government websites. The American Community Survey (ACS) data estimates for 2009, available Census 2010 data, and Franchise Tax Board income tax data provided demographic and economic analysis that focused on all fifty-eight counties of California.

The deliverables for the project consisted of a budget modeling tool, a geodatabase, and a map document that contained a standard reporting template. Also provided were maps displaying analyses of the deficit burden effects for all the counties (Figure 1.1).

Figure 1.1: Example Map Displaying Sales Tax Increase Normalized By Population.

1.3.3 Methods

The methods used for the project included assessing the data, formatting the data supplied in different file types, and meeting with the client on a weekly basis. Based on the assessment, the project data were downloaded from multiple sources including the census, ACS, and various government websites. Once retrieved, the data were re-formatted in order to be displayed in a spatial extent in Esri's ArcMap software. After formatting the data into feature classes, they were stored in a geodatabase which contained all data downloaded during the time span of the project. The data were used for analysis testing and reviewed in order to make any changes for the next implementation. Once the analysis was complete, the outputs of results were produced as maps and delivered to the client.

1.4 Audience

This report is intended for an audience that has an interest in the California economy, with a basic understanding of statistics and GIS. GIS and economic terminology are used throughout this report. Knowledge of economic data and California geography are beneficial, but not required.

1.5 Overview of the Rest of this Report

This report will provide the reader with background information, project design, implementation, results and analysis, and a conclusion of the project. Chapter 2 provides the reader with background on how government decisions are made with the assistance of GIS. Chapter 3 gives a detailed analysis of the systems used and the project design. Chapter 4 explains the conceptual and logical data models, along with data sources, collection, and preparation for implementation that contribute to the database design of the project. Chapter 5 discusses project implementation, while chapter 6 explains the analysis and results delivered by the customized modeling tool of the project. Chapter 7 completes the report with conclusions on the project and recommendations for future work.

Chapter 2 – Background and Literature Review

The California state government needs to make decisions on how to reduce the deficit while not affecting certain groups. Displaying the information spatially will make the decision process easier by being able to view the effects of raising taxes or cutting benefit programs of those groups. This chapter will examine how decision making in the government sector, understanding the geography of politics, and analyzing demographic and economic data can be achieved using a GIS.

2.1 GIS and Decision Making in the Government Sector

Using a GIS as a decision making tool in the government sector can be beneficial to both the residents of an area and its policy makers. Figure 2.1 shows a potential workflow of decision making in the government sector.

Figure 2.1: Potential Workflow of Decision Making in the Government Sector.

At a town hall meeting, local government officials discuss items on the agenda and the public can address any concerns or issues regarding those items. After the meeting, the analyst would use the input given from the public and analyze any possible effects regarding the specific issue. The analyst would submit results from the analysis for local government officials to make any policy decisions. Once a decision is made, the officials can present the results and explain how that decision was reached. This would help

present difficult information visually so the public can understand how the policy makers are deciding on solutions for the problem.

Another useful way for GIS to be incorporated into government would be for multiple departments to become familiar with the data and use them as an analysis tool to make decisions. For example, a city wanting to change a land use zone from residential to commercial could use GIS to display the zone of interest along with neighboring zones to present to the city council for discussion. Once city council members discussed all positive and negative issues with the zoning change, they could present the same information to the public. The public could voice their concerns with the change, which would help the city council in making a decision. In this case, GIS has shown ways to assist the analysis, help with visualization, manage the data, and allow for further policy development (O’Looney, 2000).

A range of GIS uses for decision making (Figure 2.2) can be performed by multiple users with different capabilities and ideas for visualization. Depending on the user, they are able to decide how they want to display the data and what the project focus should be. Using the previous example, the city planner and council members (users) were able to use high interactivity (GIS capabilities) to aid discussion with the public of the positive and negative effects of the zone change to the residents of the city. Once the users were able to present the information to the public, the visual communication of the zone change was helpful in explaining the council’s decision process to put this change into effect.

Figure 2.2: Range of GIS Uses for Decision Making (Adapted from Berry, 1994).

Government agencies use GIS to summarize large amounts of information using spatial statistic tools. Comparing feature sets in a geographic setting shows change over time, along with other factors, while deriving information to verify conclusions to help in the decision process (Mitchell, 2005). For example, a section of road going through the city has frequent traffic accidents. A transportation analyst could determine the relationship of these accidents to the speed limit on that stretch of road. If the occurrence of the accidents were clustered where the speed limit increased, this would help city officials make a decision to lower the speed limit. By analyzing the spatial relationship of these two factors, the analyst and city officials can explore why they occur and decide how to correct the issue.

Other countries use economic details to reconstruct their government. New Zealand went through an adjustment period of restructuring their economic, administrative, and social areas using a geographic restructuring model during the 1980's (Cocklin et al., 1994). The use of the model had various influences on the restructuring process on a geographic scale. It provided explanations on how human-environment relations, spatial delineation of planning the areas, and the implications of spatial differentiation during the environmental restructuring of New Zealand during the downturn of the global economy. As a conclusion to the restructuring, "with increased local autonomy in decision making and the emphasis on the 'sustainable management' paradigm, the geographical patterns of resource management decision making will warrant monitoring" (Cocklin et al., 1994, pg.466).

2.2 GIS and the Geography of Politics

Political geography is a complex subject that affects people both positively and negatively. From changing political boundaries, to reporting fraud on recovery funding projects, GIS can be used to communicate government decisions to those affected. For example, the United States government has a website that shows how awarded recovery funding is used at the state, county, congressional district, and ZIP code levels. Multiple categories are selected to map, along with details explaining the number of project funds awarded, the total amount received in a specific period, and the number of jobs created based on the projects that are to be completed (Figure 2.3).

Figure 2.3: Website of County Comparison Maps based on Education Funding and Population Under 18 (United States Government, 2011).

This is a great example of an attempt to tie policy to a location. The website is successful by displaying abundant information in an easy-to-use format. By employing choropleth maps, the data depict quantitative information in the desired area (Bolstad, 2008). Since the information is public and readily accessible, a local resident can check what projects are ongoing in their neighborhood. If the resident believes the project is mismanaged or feels that the funding is wasted, they can report fraud, waste, or abuse by submitting a complaint form to the recovery board.

2.3 The Effects of Redistricting Political Boundaries

Decennial Congressional redistricting, as called for in the U.S. Constitution, can create multiple challenges. The district boundaries critically influence the political institution's behavior and composition, which make the undertaking of defining the boundaries complex (Horn, 1999). This raises technical and theoretical challenges for all parties involved, such as trend of population changes, social, regional, and economic interest, and the proposed district area size and physical features. Gerrymandering is an example of how politicians re-draw district lines to benefit one political party. Figure 2.4 shows a result of gerrymandering in the fourth congressional district of Chicago, Illinois.

Figure 2.4: Gerrymandered Fourth Congressional District of Chicago, Illinois (Office of Senate Counsel & Research for the Minnesota Senate, 2002).

A federal district court had re-drawn the congressional district to generate a Hispanic-voting-majority while keeping the African-American votes in tact in the surrounding three districts (Office of Senate Counsel & Research for the Minnesota Senate, 2002).

Viewing socioeconomic data, existing political boundaries, and voting information in a GIS could be used in determining the new electoral district. This process is considered sensitive since there are concerns for privacy of voters, and possibilities of political gain for elected officials. Exploring all of the proposed district boundary options in a GIS would ensure careful consideration of the residents' interests and provide the best political attention to the new area.

2.4 Analyzing Demographic and Economic Conditions

Demographic and economic data are useful for analysis in many cases and can help answer the following questions:

- What is the median income for all counties in the state?
- Where are the highest unemployment rates?

- Of those counties, which have a high poverty rate?
- What are the age, gender, and ethnicity of those living in poverty?

The data used for analysis for these questions were provided by the Census of Population and Housing and help aid policy decision (Peters et al., 2004). By mailing questionnaires to all households in the United States, the United States Census Bureau is able to measure the population, housing, income, age, sex, and race characteristics every ten years (United States Census Bureau, 2011). Another survey conducted by the United States Census Bureau throughout the decade from the original census is known as the American Community Survey (ACS). The ACS releases one-year, three-year, and five-year estimates and uses three million addresses per year for the data summary (United States Census Bureau, 2008). Demographic, economic, financial, housing, and social characteristics are covered in the survey (Figure 2.5). The difference between the decennial census and the ACS is the census provides a count of the population in the United States for redistricting and congressional apportionment, whereas the ACS provides economic and social characteristics of the population (United States Census Bureau, 2008). An example of the use of this data is that an economic analyst would process the census and ACS variables to find locations of high poverty and unemployment rates in the area of interest. The analyst can provide solution options to the government to decide how to bring jobs into the area and decrease the variable's rates.

Subjects Included in the American Community Survey		
Demographic Characteristics Age Sex Hispanic Origin Race Relationship to Householder (e.g., spouse)	Social Characteristics Marital Status and Marital History* Fertility Grandparents as Caregivers Ancestry Place of Birth, Citizenship, and Year of Entry Language Spoken at Home Educational Attainment and School Enrollment Residence One Year Ago Veteran Status, Period of Military Service, and VA Service-Connected Disability Rating* Disability	Housing Characteristics Year Structure Built Units in Structure Year Moved Into Unit Rooms Bedrooms Kitchen Facilities Plumbing Facilities House Heating Fuel Telephone Service Available Farm Residence
Economic Characteristics Income Food Stamps Benefit Labor Force Status Industry, Occupation, and Class of Worker Place of Work and Journey to Work Work Status Last Year Vehicles Available Health Insurance Coverage*		Financial Characteristics Tenure (Owner/Renter) Housing Value Rent Selected Monthly Owner Costs
*Marital History, VA Service-Connected Disability Rating, and Health Insurance Coverage are new for 2008. Source: U.S. Census Bureau.		

Figure 2.5: Subjects included in the American Community Survey (United States Census Bureau, 2008).

An example of analyzing demographic and economic conditions is geography and urban studies used to understand relationships between poverty and racial economy (Wilson, 2009). A person can walk through a U.S. city and see diverse types of

economies such as real estate and service-sectors, and notice the majority of the population is either black or Latino. This type of population shows how political and economic institutions can be influenced by race and is perceived to have a negative impact.

2.5 Summary

California has the potential to reduce its deficit by making better policy and financial decisions with the aid of analysis using a GIS. GIS has been used to restructure a country, using economic data, has displayed the use of recovery funding to projects within the country, and has shown relationships between poverty and racial economic downturn. The next chapter will discuss the requirements analysis, the system design, and the project plan. The requirement analysis is necessary to help assess what the client needs before the final product was delivered. The system design will list what was needed to maintain the information for the project, while the project plan will discuss the approach for the solution to the client's problem.

Chapter 3 – Systems Analysis and Design

This chapter will address the requirements analysis, the system design, and the project plan used in solving the client's problem of identifying the primary burden per capita of the budget deficit in California. The requirements analysis section will explain the functional and non-functional requirements needed for project completion. The system design section discusses the data, software, hardware, and process components of the project. The project plan section discusses the planning involved and changes in the project schedule.

3.1 Problem Statement

Policy makers need to visualize the regional impact of the tax increases and budget cuts. The client would like to help policy makers visualize the implications of tax increases by showing areas of burden and the effects on various groups. In order for this to be achieved, a spatial budget analyst tool needed to be created with the use of available economic data. Since this is a spatial problem, identifying the primary burden per capita would assist in understanding which regions carry disproportional amount of the deficit burden, and how it could be distributed more evenly throughout the state.

3.2 Requirements Analysis

The requirements analysis was necessary to help assess what the client needed prior to project implementation. This information was determined by asking what type of functional and non-functional requirements were needed to support future analyses.

Examples of questions asked were:

- What type of output will the budget analyst tool provide?
- What type of software is needed to support the analysis?
- How will the software be maintained? Who will maintain it?
- What will the user need to know once the product is deployed?

A logic diagram shown in Figure 3.1 shows the process for the requirements analysis.

Figure 3.1: Project Logic Diagram (Tomlinson, 2007).

Creating a strategic project plan for the client would explain project goals and objectives, scope, schedule, work plan, deliverables, and a budget for the project. Once the project plan is complete, the analyst can ask, “What does the client need to know?” about his project. The questions mentioned above serve as an example of what was needed to continue with the project design. After additional questions are asked, information product descriptions (IPDs) are created to assess which required outputs for the project are necessary. Also included in the IPDs are the project workflows. This is created to be a visual guide for the client in order to view the process of the workflow. Once the IPDs are completed, acquiring the data described in the IPDs is the next step. If the data are not available, assessing whether a new data requirement is needed can be discussed with the client. If available, the data could be downloaded and formatted for input into the GIS. Once all data are retrieved, the project workflow process can begin and refer back to the IPDs if necessary.

For this project, meetings with the client were set up on a weekly basis. At these meetings, discussions about where to find project data, the progression of the project, and

other factors were mentioned to the client. Data assessments for both tax revenues were provided to the client, which led to the data preparation for analysis and the design of the geodatabase. Once the data was prepared, the client provided a budget equation to be inserted into the tool. A Python script was written to incorporate the equation and a model was built to include the script. Using the data from the project, the tool was tested with multiple inputs. The output results were shown to the client and discussed any changes that needed to be performed prior to delivery of the tool.

3.2.1 Functional Requirements

Functional requirements for the project were as follows:

- System must include a budget analyst tool for counties and ZIP codes.
- System must have a geodatabase with sales and income tax data.
- System must display potential sales and income tax increase data results from budget analyst tool on individual choropleth maps for comparison.
- The user must be able to adjust percentages of sales and income tax increases.
- System must have a standardized symbology for the budget analyst tool output.
- System must have a standard reporting template that produces maps.

These functional requirements were used as a foundation for the project's system design.

The budget analyst tool was built with an equation the client had provided. A Python script was written to include the equation and a model was produced to accommodate the sales and income tax data.

A geodatabase was created to serve as a warehouse for the sales and income tax data used throughout the project. The California state budget was used as a conceptual model for building the geodatabase. The feature datasets stored in the geodatabase were separated as *county* and *ZIP code*. In these feature datasets, feature classes of the sales and income taxes were stored based on their region.

The choropleth maps were produced for analysis of potential sales and income tax increase results that showed which regions were affected. These maps were to aid the client in visualizing alternatives on how to distribute of the financial burden equally throughout the state. The data used for the maps were imported from multiple government websites, ACS, and comprehensive annual financial reports (CAFR) from each county in California. State sales and income tax values by county and ZIP code were formatted for analysis using Esri's ArcMap. This helped to display the information spatially, query for desired items, and perform other functions with ease. Basemap data (such as streets, county, and city boundaries) were used to select the areas of interest.

The user would use the budget analyst tool created to adjust the percentage of tax increase and display the financial burden throughout the state results as a new output feature class. The data were taken from the geodatabase and processed through the model, which created an output for display in a map document. The user would be able to interpret the data by comparing both taxes for analysis.

A symbology standard for the choropleth maps would be provided in order to keep all new layer outputs consistent when comparing analysis. Once the maps were created for all California, the user would be able to determine if the tax increase had a spatially random pattern or if the data were clustered to certain areas with negative effects using spatial statistic tools provided in ArcMap.

Once the user is satisfied with the results, the standard reporting template (such as a map document) would be used to print results with minimal cartographic elements. This would help the user spend less time on the cartographic design of the maps and print the maps quickly for analysis. The template would have the ability to be altered to meet future map document specific requirements.

3.2.2 Non-Functional Requirements

Non-functional requirements consist of technical, operational, and transitional requirements that contain limitations on the client's working environment. The client would need to determine if and how to accommodate the project's non-functional requirements before completion.

A software technical requirement was to have consistent data transformations to and from non-spatial and spatial outlets. Downloaded data from various sources can have different file extensions, therefore the data would need to be converted to a dBase (.dbf), or Excel file (.xlsx) format and loaded into the project geodatabase. In order to view the data in ArcMap, the tabular data would be joined to a feature class that contains spatial coordinates and would be converted into a vector format. This would ensure that the data would be available for importing into a geodatabase and store all project related data.

Operational requirements needed for the project to run efficiently were for the user's computer to have ArcGIS 10.0 software installed and the computer must meet the software's requirements. The budget analyst tool would only work on ArcGIS software since the tool was created inside of ArcCatalog and executed inside a map document. The hardware of the user's computer would need to meet the requirements of ArcMap to ensure system performance and responsiveness expectations.

Transitional requirements were to provide all the data, modeling tool, map documents, products, and necessary documentation to the client. The data stored in the map documents and geodatabases contained completed metadata where applicable.

A summary of the non-functional requirements is found in Table 1.

Table 1. Non-Functional Requirements

Domain	Requirement
Software	Consistent data transformations to and from non-spatial and spatial outlets.
Operational	Must use Esri ArcGIS 10.0 or higher software.
	Hardware must meet ArcGIS requirements.
Transitional	Provide geodatabase containing data.
	Provide budget analyst tool.
	Provide map document and standard reporting template.
	Provide any necessary documentation.

3.3 System Design

The budget analyst modeling tool was designed by the functional and non-functional requirements. The project focused on building a geodatabase, a modeling tool with a user-friendly interface, and standard reporting template that would produce multiple regional maps for analysis. In order to create these products, the process, hardware, software, and data components were necessary for completion (Figure 3.2).

Figure 3.2: System Design Components.

3.3.1 Process Component

The use of Esri's ModelBuilder in creating the budget analyst modeling tool was essential. The output feature class produced by the tool was major component to the

project. Once the output was added to the map document, the user would be able to display and print the results using the standard reporting template.

3.3.2 Hardware Component

The GIS was built by using a Dell Precision M4500 laptop computer with an Intel® Core™ i5 central processing unit with 8 gigabytes of RAM, 250 gigabytes of storage, and a CD/DVD drive with writing ability. All files for the project were stored on a terabyte external hard drive and were backed up weekly. High-speed internet provided access to data from the client's server. The server requirements needed for this project were to have accessibility, encryption, hosting, and an environment for the system operation and disaster recovery. The user would have access to the project data and stored in an accessible server. By storing the data in the server, the guidelines, regulations, and constraints set up by the user or information technology (IT) technician would ensure that the data would maintain consistency and functionality. If a system disaster were to occur, the IT technician would be able to recover any data lost by retrieving available archives. A reliable and speedy connection to the intranet and internet would be necessary for access to data from the server and downloading data from websites. An encryption could be set up by the user or project manager in case other members of the project team would like to view the data, but are not allowed to write or edit the data depending on the sensitivity of the project.

3.3.3 Software Component

The budget analyst modeling tool was built specifically from ArcGIS Desktop version 10.0 software. This was useful to the client since he also uses the software as his preferred analysis tool. The modeling tool consisted of data management tools which included adding, calculating, and joining fields. Also included was the Moran's I spatial statistic that included a HyperText Markup Language (HTML) report to analyze the results. The tool interface provided helps the user enter the requested parameters and runs the tool for quick and easy analysis.

In order for the tool to run efficiently, all software updates were recommended for installation in a timely manner. Esri announces software updates to ArcMap by email to those who have purchased the software. If the software is not kept up to date, this could affect the project functionality. Ensuring software updates were installed for ArcMap would help patch any discrepancies and enhance the software.

3.3.4 Data Component

The data used for the project came from various sources. Data downloaded from the Franchise Tax Board (FTB), U.S. Department of Commerce Bureau of Economic Analysis (BEA), California Board of Equalization (BOE), and Internal Revenue Service (IRS) were personal income tax and sales tax statistics for 2009-2010 fiscal year. Other data included ACS income data estimated for 2009, and Esri's Business Analyst demographic and income estimates for 2008. The data were in Excel or Portable Document Format (PDF) file formats and were converted to a tabular format. Once the data were normalized, the tables were joined to similar fields in existing feature classes

and were exported as new individual feature classes. The data were loaded into a geodatabase for common georeference, which will be discussed in Chapter Four.

3.4 Project Plan

At the start of the project, a plan was developed as a guide to ensure the project met the client's goals and was executed in a timely manner. A schedule was created at the start of the project to stay on course with deliverables. However, as the year continued, the schedule was revised to incorporate delays (Figure 3.3).

Figure 3.3: Original and Revised Project Schedules.

The project began with reviewing the 2009 state budget revenues and expenditures. A data assessment was created for the expenditures to view the impact of the financial adjustments made on certain groups. Once completed, the client requested the data assessment that led to the decision of which policies would be the project focus. A decision to select the sales and income tax revenues was supported by the client since both revenues combined made up approximately 72% of the state budget revenue. Since the client only provided one dataset for tax revenues, a search for the data was needed.

The data were downloaded from multiple government sources and were prepared for scrubbing and loading into the GIS. The geodatabase was designed by using the state budget as a conceptual model. Once the geodatabase was completed, the data were imported by creating a join on fields in existing feature classes related to the project. The data were exported as new separate feature classes and stored in the geodatabase. The budget analyst tool was created using a Python script to integrate the client's equation. Also included was a model containing the data management tools to incorporate the project data. A user interface was developed to provide the ability of adjusting tax percentages to produce output results. The tool went through multiple iterations, providing outputs for analysis. A standard reporting template was created for ease of printing the results for comparison analysis. All items included in the project were provided to the client upon completion.

The original project plan was to analyze all of California's general fund revenue and expenditure sources. Since all of the sources can be affected based on the financial adjustments of the budget, the plan was to view which programs are heavily funded and how that affects people receiving support from that program. After multiple data searches turned up no data for most of the sources, contacting the Governor's office would have been the next step in the data search. While it was unknown how long communication between the client and state would take, time constraints prevented that process from happening, since the project deliverables were due mid-December 2011. Missing data and time constraints were the main setbacks to the original project plan. It was decided to select two sources from either the general fund revenues or expenditures listed in the state budget as the focus for the analysis. Using the modeling tool provided as a deliverable, the client could choose to apply the tool to all other sources and make modifications for those additional analyses.

3.5 Summary

With the use of a project plan, requirements analysis, and system design, the project was able to start with a goal to solve the client's problem. The requirements analysis helped assess what type of functional and non-functional requirements were needed for the project to function. The system design used data downloaded from multiple sources, and created a geodatabase and modeling tool to display and organize the data stored on the client's server. Although the project plan had schedule changes, the project was able to finish and provide the client with the deliverables as requested.

Chapter 4 – Database Design

This chapter discusses the process in creating the conceptual and logical data models of the project geodatabase. The conceptual data model exhibits similar characteristics from the state budget, whereas the logical data model used data downloaded from multiple government agencies. Also discussed in this chapter are the list of data sources, and loading and scrubbing of the data downloaded.

4.1 Conceptual Data Model

The design of the conceptual data model was based on the working components of the state budget. Since the state budget is made up of multiple *expenditures* and *revenues*, the focus was aimed toward income and sales taxes (Figure 4.1), which account for approximately 72%, of the state's revenue every year.

Figure 4.1: California State Budget Conceptual Model.

Income tax revenue is generated by imposing a tax on individual income. Sales tax revenue is generated by taxing items for sale at the time of purchase. The statewide sales tax rate of 7.25% is the minimum tax to be charged on items; however, a county may add a district tax to the statewide rate percentage, which varies from 0% to 1.00% (California State Board of Equalization, 2011). The tax-assessed revenue is allocated to the counties

from the state and is further distributed to cities. Using these revenues, the counties are able to fund their local budget.

4.2 Logical Data Model

The logical data model was built in ArcGIS Desktop using ArcCatalog to create the California state budget file geodatabase. The geodatabase stored financial data related to income and sales tax revenue, personal income summaries, and population counts for 2009 on county and ZIP code levels. The county and ZIP code feature datasets (Table 1) were used to organize the related feature classes produced for the project.

Table 2. California State Budget Logical Data Model

Feature Dataset	Feature Class	Description
County	CA_BEA_CountyPerCapitaPersonalIncome2009	Contains county name, and per capita personal income statistics for 2009.
	CA_BEA_CountyPersonalIncome2009	Contains county name, and personal income statistics for 2009.
	CA_BEA_CountyPopulation2009	Contains county name, and population statistics for 2009.
	CA_BOE_CountySalesTaxRevenueFY0910	Contains county name, third quarter 2009 through second quarter 2010 taxable sales, total 2009-2010 fiscal year taxable sales revenue, state sales tax rate, and total state sales tax revenue for 2009-2010 fiscal year.
	CA_FTB_CountyPersonalIncomeTax2009	Contains county name, population statistics from July 2009, number of returns, adjusted gross income, tax assessed, tax assessed per capita, and working population statistics.
	CACounties	Contains county name and demographic statistics.
	CountyIncomeMadeandAGI	Contains all data joined between CA_BEA_CountyPerCapitaPersonalIncome2009 and CA_FTB_CountyPersonalIncomeTax2009 along with total tax revenue as a percentage of income by county.
ZIP Code	ZipCode2009	Contains ZIP code, post office name (city), and demographic statistics.
	ZipCodeIncome_Made	Contains all data joined between ZipCode2009 along with number of returns, adjusted gross income total tax liability, tax assessed per capita, and adjusted gross income per capita.
	ZipCodeSalesRevenueMade	Contains all data joined between ZipCode2009 along with taxable sales for 2007, 2009, 2010; the sum of taxable revenue for both 2009-2010 calendar years, the total taxable revenue for 2009-2010 fiscal year, state sales tax rate, state sales tax revenue, state sales tax revenue per capita, and total tax revenue as a percentage of income by ZIP code.

Using a feature dataset ensures that all data-related feature classes have the same spatial reference when the dataset is created. Since the data will be shared with other users, this removes the possibility of displaying the data incorrectly by having them built into the datasets. The county feature dataset is stored in the geodatabase and included several feature classes, such as California county boundaries, county sales tax revenue for 2009-2010, county population counts for 2009, and county income.

The ZIP code feature dataset also contained boundaries, tax information, and income statistics, which were used for analysis. The boundaries for both feature datasets displayed spatial extents of the county and ZIP code areas. The 2009-2010 fiscal year sales tax revenue feature class displayed total taxable sales for 2009-2010 by area and the state sales tax rate, which were multiplied to create the state sales tax revenue field. The 2009 population counts feature class showed population counts per area for 2009. The county income statistics feature class included fields such as personal income data, personal income per capita data, adjusted gross income, median income, and tax assessed per capita in 2009 for all areas.

4.3 Data Sources

The data collected for this project were from multiple government agencies that included the U.S. Bureau of Economic Analysis (BEA), California Board of Equalization (BOE), Internal Revenue Service (IRS), and State of California Franchise Tax Board (FTB). These sources contain complete datasets in regards to financial statistics needed for the analysis. Other data that were downloaded but not used in the analysis were from the American Community Survey (ACS) and county comprehensive annual financial report (CAFRs). The ACS data were difficult to navigate during the selection and download process, and once retrieved were not the results desired for analysis. County CAFRs were downloaded from individual county websites, but were not consistent with the information presented among other CAFRs. There were a couple of CAFRs that were not prepared for the same 2009-2010 fiscal year needed for the project, or the county did not upload the CAFR to the county website for public download. These datasets were incomplete and of no use to the analysis; therefore a decision was made to discard the data.

4.4 Data Scrubbing and Loading

Since the client provided taxable sales data for ZIP codes, the suggested list of government agencies were recommended as good sources to collect the rest of the data (Figure 4.2).

Figure 4.2: Data Retrieval Method.

Using the BEA website, a search for individual income tax statistics provided downloadable personal income, population, and per capita personal income information for 2009 by county as a comma separated values file (.csv). The IRS website was used to download individual income tax return data by county and ZIP code for the tax year 2009 in separate Microsoft Excel files (.xls). The file included attributes such as number of returns, number of exemptions, adjusted gross income, and wages and salaries. Taxable sales by county and ZIP code for third and fourth quarters of 2009, and first and second quarters of 2010 were retrieved from the BOE website as separate Excel files. All four quarters were added together to simulate 2009-2010 fiscal year data for analysis and were saved as a new Excel file. The FTB website provided data containing population count from July 2009, the number of returns made in 2009, adjusted gross income, median income, and tax assessed by county in an excel file, and ZIP code in an PDF file.

The files acquired from the data sources were stored in a geodatabase. All files were saved as Microsoft Excel 2010 files (.xlsx) and with an extension of “_esri” in the file name. This ensured that the original data downloaded remained intact during transformation and loading into the geodatabase of the new files. If the new file was corrupt or needed data verification, the original files were available for review or recreation of the new file. Other transformations included column heading spaces removed, or headings that started with numbers required text preceding the number. Naming conventions were also addressed since the data had no spatial reference included and required a join on the county name or ZIP code. For example, on the IRS file, the county name included “County” after the name (ex. Alameda County). Since the file was joined to the county basemap feature class, “County” was removed from all the rows so that only the name (Alameda) would be able to join the files. A majority of the files contained a number format for columns containing number amounts. However, some columns had dollar amounts as text and needed to be formatted as a number. Notes that

were included in the files were also removed. Once these files were scrubbed, the Excel files were imported into the geodatabase as tables and were brought into a map document. In the map document, the county basemap feature class was used as a base for joining all tables without spatial reference by county name or ZIP code. Once the join was successful, the data were exported as new feature classes and stored in its specified feature dataset located in the geodatabase. After this step was completed for all the files, the data was ready for analysis and tool execution.

The budget analyst tool (discussed in Chapter 5) was stored in the geodatabase. The tool used the county and ZIP code feature dataset's inputs and operated within the map document with an interface that requested the sales tax percentage from the user to apply to the state sales tax revenue. During the tool's execution, the income tax percentage recommended was derived in order to maintain the state's revenue from the income and sales taxes. An output feature class showing the state redistribution of tax percentages was created and stored in the geodatabase, and automatically added to the map document for further analysis.

4.5 Summary

The California state budget geodatabase was designed using a conceptual data model based on the working components of the state budget. The logical data model was built within ArcGIS and stored data downloaded from the BOE, BEA, IRS, and FTB. Also included within the geodatabase was the budget analyst tool which used the data for analysis within the map document. The retrieved data needed scrubbing in order to be loaded into the geodatabase. Once the data were transformed into individual feature classes within datasets, the implementation process could begin, as described in the next chapter.

Chapter 5 – Implementation

This chapter discusses the implementation of the project data, model, and script tool. Also discussed are how the data were explored, what type of conversion the downloaded files needed, and how the data were refined for input into ArcGIS.

5.1 State Budget Data Exploration

Searching for sales and income data was a priority since the client did not provide most of the data. A search for taxable sales in California led to the California State Board of Equalization (BOE) website that provided a range of years to select the data needed. The data were available for download by calendar year and by quarters as Excel or PDF files. Since the California State Budget is conducted on a fiscal year calendar, which starts on July 1st and ends on June 30th on the following year, the taxable sales data were downloaded as Excel files for the third and fourth quarters of 2009, and first and second quarters of 2010 by county. This would ensure the data would encompass all data for the fiscal year of 2009-2010. Included in the downloaded table were parameters such as county name, sales tax permits totals, and taxable transaction totals.

Taxable sales data for ZIP codes was the only dataset provided by the client. The data were purchased from the California Retail Survey website and were available in an Excel format. The survey included retail results on all counties and cities with sales data from different types of merchandise stores. Also included were sales figures for 2000 through 2010, along with changes in growth percentages. The client formatted the data for his personal use and provided the final format for analysis that only included total taxable sales for calendar years 2007, 2009, and 2010 by ZIP code.

The income tax statistics data were found by searching the Franchise Tax Board (FTB) website. The county data were provided in an Excel spreadsheet format and the parameters included county name, population, number of returns, adjustable gross income, median income and rank, number of joint returns, and tax assessed for 2009. For ZIP code income tax statistics, another search was performed on the FTB website that yielded a PDF containing ZIP codes, number of returns, adjusted gross income, and total tax liability parameters for the 2009 calendar year.

A search on the U.S. Department of Commerce Bureau of Economic Analysis provided additional income statistics by county dated from 1969 through 2009. Parameters included in the downloaded comma separated values (.csv) file were personal income, population, and per capita personal income and the federal information processing standard code for each county.

5.2 Data Conversion

Converting the downloaded data from various formats for use in ArcGIS was required. Since most of the data were in Excel or .csv formats, the files needed minor refinements before input into ArcGIS and are mentioned in the next section.

The ZIP code income tax statistics data from the FTB website were downloaded as a PDF, and were converted to an Excel file with Able2Extract Professional software provided by the client (Figure 5.1).

Franchise Tax Board Personal Income Tax Statistics by Postal ZIP Code 2009 Calendar Year							
Zip Code	Returns	Adjusted Gross Income (\$000)	Total Tax Liability (\$000)	Zip Code	Returns	Adjusted Gross Income (\$000)	Total Tax Liability (\$000)
90001	18,722	466,563	3,693	90047	20,693	703,883	10,124
90002	15,730	386,238	2,419	90048	12,847	445,164	63,742
90003	21,493	509,237	3,133	90049	19,507	4,817,698	387,445

Franchise Tax Board			
Personal Income Tax Statistics by Postal ZIP Code			
2009 Calendar Year			
		Adjusted Gross	Total Tax
Zip Code	Returns	Income (\$000)	Liability (\$000)
90001	18,722	466,563	3,693
90002	15,730	386,238	2,419
90003	21,493	509,237	3,133

Figure 5.1: Franchise Tax Board ZIP Code Income Tax Statistics as a PDF with Data Converted to Excel Format Using Able2Extract Professional Software.

Using Able2Extract Professional, the user was able to navigate to the PDF location saved and open it using the software. Once opened, an option is given to select all of the pages in the PDF or select an area to be converted. When the final area is selected, the user can choose what file type to convert the selection. Since an Excel file type was preferred, an option was given to create an automatic Excel output with a general table structure or select a custom format which specifies a column structure inputted by the user. After an automatic or custom structure is decided upon, the software converts the file and opens in Excel for viewing and the user can save the file to a specified directory.

5.3 Data Refinement

All downloaded files were refined and organized to simplify names, remove spaces from column headings, change any numerical data from text to numbers, or convert dollar amounts from thousands of dollars to actual amounts. An example shown in Figure 5.2 displays how the BOE downloaded county sales data column headings have merged cells, spacing in column names, and a title preceding the column names before re-formatting.

TABLE 2. TAXABLE SALES, BY COUNTY, FOURTH QUARTER 2009 (Taxable transactions in thousands of dollars)								
County	Sales tax permits as of January 1, 2010			Taxable transactions				
	Incorporated cities	Outside incorporated cities	Total	Incorporated cities	Outside incorporated cities	Unallocated b/	Total	Percent change year to year
Alameda	36,783	1,886	38,669	\$4,728,911	\$152,123	\$632,478	\$5,513,512	-5.1
Alpine	--	66	66	--	2,276	2,686	4,962	0.4
Amador	656	697	1,353	25,975	57,459	10,353	93,786	-5.8
Butte	4,209	1,525	5,734	495,749	51,661	65,647	613,058	-4.8
Calaveras	179	1,287	1,466	12,401	41,898	10,752	65,051	-8.1
Colusa	276	217	493	33,754	39,053	16,345	89,153	18.2

County	Total3q09TaxableSales	Total4q09TaxableSales	Total1q10TaxableSales	Total2q10TaxableSales
Alameda	5,148,605,353.00	5,513,512,084.00	4,917,425,563.00	5,308,792,776.00
Alpine	4,951,886.00	4,961,936.00	7,796,178.00	3,943,449.00
Amador	87,852,671.00	93,786,361.00	80,237,463.00	90,606,731.00
Butte	597,951,551.00	613,057,624.00	553,698,392.00	621,483,887.00
Calaveras	71,421,988.00	65,050,623.00	57,322,075.00	69,471,661.00
Colusa	117,774,370.00	89,152,551.00	58,902,191.00	115,358,777.00

Figure 5.2: California State Board of Equalization Data as Downloaded with Data Reformatted for Import into ArcGIS.

Although other parameters were given, the only desired columns, *County* and *Total Taxable Transactions*, were stripped out and saved as a new file, *CountySalesTaxRevFY2009-10.xlsx*. The *Total Taxable Transactions* dollar amounts were in thousands of dollars and were changed to show the true dollar amount by multiplying the data in the column by 1000. Doing so would ensure that all dollar amounts shown in the project data were consistent. Since there were four quarter years' worth of data, the *Total Taxable Transactions* columns on the remainder quarters were extracted and added to the *CountySalesTaxRevFY2009-10.xlsx* file. Once all the data were input into one file, additional columns included were *TotalFY0910TaxableSalesRevenue*, *StateSalesTaxPercentage*, and *TotalStateSalesTaxRevenue*. Since the total taxable sales revenue was needed for the 2009-2010 fiscal year, the third and fourth quarters of 2009 and the first and second quarters of 2010 were added together in the *TotalFY0910TaxableSalesRevenue* column. The *StateSalesTaxPercentage* column included the value of 0.0725, and was used to calculate the total state sales tax revenue by multiplying the *TotalFY0910TaxableSalesRevenue* and *StateSalesTaxPercentage* columns. The calculated value was displayed in the *TotalStateSalesTaxRevenue* column for all fifty-eight counties.

The BEA downloaded county income data shown in Figure 5.3 was separated in three different files for analysis: Personal income, population, and per capita personal income.

FIPS	Region	Table	Line code	First year	Line Title	Area Name	1969
6000	8	CA1-3	1	1969	Personal income	California state total	89336977
6000	8	CA1-3	2	1969	Population	California state total	19711000
6000	8	CA1-3	3	1969	Per capita personal income	California state total	4532
6001	8	CA1-3	1	1969	Personal income	Alameda, CA	4841528
6001	8	CA1-3	2	1969	Population	Alameda, CA	1060099
6001	8	CA1-3	3	1969	Per capita personal income	Alameda, CA	4567

FIPS	LineTitle	AreaName	TaxYear1969	TaxYear1970	TaxYear1971	TaxYear1972
06000	Per capita personal income	California state total	4532	4801	5027	5451
06001	Per capita personal income	Alameda, CA	4567	4875	5117	5480
06003	Per capita personal income	Alpine, CA	4744	4163	4231	4561
06005	Per capita personal income	Amador, CA	4097	4222	4397	4818
06007	Per capita personal income	Butte, CA	3402	3624	3810	3971
06009	Per capita personal income	Calaveras, CA	3530	3841	3980	4452
06011	Per capita personal income	Colusa, CA	5776	5628	5859	6648

Figure 5.3: U.S. Bureau of Economic Analysis Data as Downloaded with Per Capita Personal Income Data Reformatted for Import into ArcGIS.

Prior to separation of the three categories, the column headings needed spaces removed from the text. The year column titles needed the text *TaxYear* in front of the number before importing the files into the geodatabase since ArcGIS does not allow columns to start with numerical values. Once the reformatting was completed, the file containing all three categories was imported into the geodatabase and joined to the CACounties feature class. Queries were created in the joined file to export a new feature class for each category and were named accordingly:

1. *CA_BEA_CountyPerCapitaPersonalIncome2009*
2. *CA_BEA_CountyPersonalIncome2009*
3. *CA_BEA_CountyPopulation2009*.

The ZIP code data downloaded from the FTB and California Retail Survey needed minor refinement before importing into the geodatabase. For example, the California Retail Survey data did not have any spaces in the column headings but needed the sales tax revenue calculated prior to analysis. Since the data provided taxable sale totals for 2009 and 2010 calendar years, a sum of both years was taken and divided by 2 to receive an estimate 2009-2010 fiscal year. In a new column labeled *SalesTaxRevbyZip*, the fiscal year estimates were multiplied by the state sales tax 0.0725 which equaled the sales tax revenue for the state. The data were imported into the geodatabase, joined to the ZIP code feature class from Business Analyst, and exported as a new feature class, *ZipCodeSalesTax2009*. For the FTB ZIP code income data, spaces were removed from the column headings, and the *TotalTaxLiabilityinThou* column containing the dollar amounts in thousands were calculated in a new field named *TotalTaxLiability* by multiply it by 1000. The file was also imported into a geodatabase, joined to the ZIP code feature class and exported as a new feature class, *ZipCodeIncome_Made*.

5.4 Data Model

The data model is considered the core of the project that supports the client with his analysis. Figure 5.4 shows the general flow of the California budget analyst tool.

Figure 5.4: General Flow of County and ZIP Code Data Models.

The California budget analyst tool is executed first since the rest of the model requires the output values in order to run successfully. When the county or ZIP code model is selected, a pop-up window is displayed and asks the user to input a desired sales tax percentage for analysis (Figure 5.5).

Figure 5.5: Pop-Up Window Display for User Sales Tax Percentage Input into the Model.

Once the user clicks “OK”, the Budget Analyst Tool runs through a python script that is connected to the model. The sales tax percentage entered is represented as a percentage and is divided by 100 to represent the fraction. This number is ready to run through the equation provided by the client (Figure 5.6):

Figure 5.6: Budget Analysis Equation Provided by the Client.

The recommended sales tax percentage entered by the user is multiplied by the sales tax revenue and added to derived income tax percentage. Then the product is multiplied by the average taxable income rate multiplied by the total taxable income per county. The \$72 billion is the sum of the income tax revenue and sales tax revenue taken from the 2009-2010 fiscal year state budget in which the equation is equal to. By performing algebra to derive the recommended income tax percentage, the equation was altered as the following (Figure 5.7):

Figure 5.7: Adjusted Equation Used in the Data Model.

Once the recommended income tax is derived, the rest of the model is able to run. The model takes the input sales tax feature class, copies all the attributes and creates a new feature class with the naming extension *_test*. The *Sales Tax Calculation* and *Recommended Sales Tax* percentage fields are added to the new feature class. In the *Recommended Sales Tax* field, the sales tax percentage entered by the user is divided by 100 and is populated for all fifty-eight counties. The *Sales Tax Calculation* field takes the *Recommended Sales Tax* and multiplies it with the *State Sales Tax Revenue*, which would satisfy *A * Sales Tax Revenue* portion of the equation. The input income tax feature class is joined with the sales tax feature class and adds the *Income Tax Calculation* and *Recommended Income Tax* fields to the *_test* feature class. The *Recommended Income Tax* field is calculated by using the derived income tax percentage output provided by the budget analyst tool and is divided by 100. Once the income tax percentage is populated, the *Income Tax Calculation* field calculates the *Average Taxable Income Rate* multiplied

by *Total Taxable Income per County*, then multiplied by the *Recommended Income Tax*. The *State Income* field is added and calculated by adding the *Income Tax Calculation* and *Sales Tax Calculation*, which equals ~\$72 billion when all fifty-eight counties are summed together. The last step of the model is to perform the Moran's I tool which uses the feature locations and attribute values to measure if the results are spatially auto correlated. A report of the results would be accessible from the results window and stored locally on the users' computer for further analysis (Figure 5.8).

Figure 5.8: Tool Output Accessibility from the Results Window (Esri, 2010).

The report shows the Moran's I and p value, z-score, along with a bell curve depicting a dispersed, random, or clustered pattern of the data (Figure 5.9).

Figure 5.9: Spatial Autocorrelation HyperText Markup Language (HTML) Report (Esri, 2011).

After the spatial autocorrelation tool has executed, the feature class is added to the map document and is ready for analysis. Once the user has viewed the information relating to the report, the user would choose to re-run the model with a different sales tax percentage to yield different results.

5.5 Summary

The implementation stages of this project were to explore, convert, refine, and model the data. The exploration of the state budget data was found from multiple government source which included the BEA, BOE, FTB, IRS, while the California Retail Survey was provided by the client. Most files were provided in a tabular format, while one dataset downloaded was converted from a PDF to an Excel document. The refinement of the data

was minor and mainly needed cosmetic changes to column headings. Once the data were imported into the geodatabase, the model was prepared to run with the necessary feature classes and produce results for display and analysis in a map document. The next chapter will discuss the budget analyst tool's implementation results.

Chapter 6 – Results and Analysis

This chapter explains the results and analysis of the budget analyst tool outputs. The primary goal of this project was to provide the client with a tool to perform analyses of the sales and income tax revenues in the California state budget, along with a geodatabase containing the data to run the tool.

6.1 Budget Analyst Tool County Output Results

Once the budget analyst tool has run, an output feature class was produced and added to the map document (Figure 6.1). The output feature class contained all the additional fields and calculations by the model mentioned in Chapter 5.

Figure 6.1: Budget Analyst Tool County Output Feature Class.

Originally, a symbology layer was to be applied to the feature class when it was added to the map document. Instead, the output feature class was interpreted and displayed by the

user based upon the parameters analysis. Adding a symbology layer would have restricted the user temporarily from viewing the data by giving a customized portrayal as a first impression.

After the output feature class has been added to the map document, the user can begin the analysis. The sales tax used in this scenario was 8.25% and was used to calculate county-level revenue from sales. To assess the relative tax revenue, the *SalesTaxCalculation* results were normalized by *Population070109* with a standard deviation classification ranging from *Far Below Average* to *Far Above Average* (Figure 6.2).

Figure 6.2: User Input Sales Tax Rate Revenue Results Shown in Map Template.

The counties shown in red are areas in which people spend the least per capita on taxable goods and services. Counties shown in yellow are areas that generate an average amount

of per capita revenue for the state. The counties shown in dark green generate the far above average per capita sales tax revenue for the state.

Using the same classification to display the recommended income tax rate derived from the budget analyst tool, Figure 6.3 shows which areas would create the most revenue for the state.

Figure 6.3: Derived Income Tax Rate Revenue Results Shown in Map Template.

Twenty-five out of fifty-eight counties are below average and do not generate a significant amount of per capita revenue for the state. Twenty-two counties are average in producing per capita income tax revenue, and only four counties shown in dark green produce the most per capita revenue for the state. The far above average counties have the highest income made per capita, which explains how most of the income tax per capita revenue is generated in these areas.

The sales and income tax rates maps were classified with user-defined symbology and the client was able to see the information as a side-by-side comparison. Figure 6.4 shows the sales and income tax rate maps combined as a bivariate map to show what areas were most and least significantly by changes in both sales and income taxes.

Figure 6.4: Bivariate Map of Sales and Income Tax Per Capita Revenue Results.

The bivariate map displays the areas that have both low sales and income tax revenue, areas that have both high tax revenues, and all other combinations in-between. For instance, the San Francisco Bay counties shown in dark green presented sales and income tax revenue are high in these areas. This could be attributed to higher personal income in these counties and higher levels of taxable spending. Counties shown in the low end of both tax revenues reveal that lower personal income groups do not spend money other than buying essential items. Although not included in the map template, viewing the

results as a bivariate map could help the client make assumptions about who is affected by the tax increases.

6.2 Budget Analyst Tool ZIP Code Output Results

Once the user has run the budget analyst tool for the county data, the user can choose to run a separate tool that can be applied to the ZIP code data. Both the county and ZIP code budget analyst tools use similar Python scripting (Appendices A and B), but call on different data to produce results. The output feature class shown in Figure 6.5 contained all the additional fields and calculations needed by the model mentioned in Chapter 5.

Figure 6.5: Budget Analyst Tool ZIP Code Output Feature Class.

After the output feature class has been added to the map document, the user would likely notice that there are ZIP codes missing. This is because the Franchise Tax Board (FTB) ZIP code data contained approximately 2400 ZIP codes for California. Since the FTB data had no spatial data associated with it, a spatial join using Esri's Business Analyst ZIP code feature class was necessary. Esri's feature class contained approximately 1700 records, and when joined with the FTB tabular data, approximately 1600 ZIP codes are displayed (Figure 6.6).

Figure 6.6: Venn Diagram of Approximate Data Matched Using Esri ZIP Code and Franchise Tax Board ZIP Code Datasets.

Thirty-four percent of the data were not matched through the spatial join because most cities contain Post Office (P.O.) boxes that are associated with their own ZIP code. The FTB data had contained ZIP codes of P.O. boxes, whereas the Esri feature class only included ZIP codes with actual land area. As a test, the FTB ZIP code data were geocoded in ArcMap and a match was produced for all ZIP codes. As a further inquiry, the City of Redlands was chosen to view the number of different ZIP codes associated. While three ZIP codes are related to Redlands, it was noted that two ZIP codes were actual land area of Redlands; one ZIP code was associated with a P.O. Box. The FTB data contained all three ZIP codes, whereas the Esri feature class only included two ZIP codes. Other ZIP codes included in the Esri feature class that were not found in the FTB data contained *000##* ZIP codes that were designated by Esri to state national parks. These ZIP codes were searched for using the United States Postal Service website and found that the *000##* ZIP codes did not exist.

Using the available ZIP code data from the output feature class, the user can begin the analysis. The sales tax used in this scenario was 8.25% and was calculated to show how much money is generated by the state at that rate. The *SalesTaxCalculation* results were normalized by *POP2009* with a Natural Breaks (Jenks) classification (Figure 6.7).

Figure 6.7: User Input Sales Tax Rate Revenue Results at ZIP Code Level.

During the classification process, the data displayed multiple areas with no sales revenue and no population, which were marked as insufficient data. Since the state has approximately 45% of government owned land, a majority of that land is national parks in which no one is allowed to live. A query was applied to the data to show areas with population greater than 200 persons and sales tax revenue greater than \$200. Although various numbers were tried in this query to further eliminate the larger outliers, the numbers listed above only apply to this scenario.

Figure 6.7 showed most of the ZIP codes generate a lower amount of per capita revenue for the state. Some of the areas that produced a higher amount of sales revenue per capita were areas that also produced a higher amount of income revenue per capita (Figure 6.8).

Figure 6.8: Derived Income Tax Rate Revenue Results at ZIP Code Level.

An example of where both sales and income tax per capita revenues generate a significant amount for the state would be the City of San Francisco. Figure 6.9 shows one of the ZIP codes associated with San Francisco that would make a large amount of sales revenue per capita if a 1% increase is applied.

Figure 6.9: A ZIP Code Area of San Francisco Shown with Business Locations.

The majority of the business shown in the map are retail, restaurant, and car dealerships, which have the most sales volume according to Esri's Business Analyst data. An assumption would be that since there were approximately 26,000 people in this ZIP code in 2009, a majority of the per capita revenue earned could be attributed to the services provided in the area. Another assumption would be that since this ZIP code lies within a major intersection of Highway 101 and Interstate 80, motorists that pass through these highways may be more likely to stop and make purchases from these businesses since they are clustered together. Additional analyses could be performed with this data as to what industries are generating the most revenue for the state. This is described further in Chapter 7.

6.3 Summary

The sales and income tax revenue maps shown in this chapter were an example of the analyses that can be performed with the calculated values provided with the budget analyst tool. Tax revenues were found to be substantial in urban places, therefore if there was an increase in the tax rate, the affect would be on a large number of people living in the area. The client has the ability to make assumptions as how to spread tax increases evenly throughout the state by running the model with different tax rates until the desired

effect is achieved. This was useful to the client because he was able to view the effects of the tax increases and understand the spatial pattern associated with the increases.

Chapter 7 – Summary and Future Work

This final chapter concludes and summarizes the creation of the budget analyst tool for the client and presents potential ideas for future work. The project fulfilled five out of six functional requirements and satisfied all non-functional requirements established by the client, and the tool is ready for use once transferred to the client.

7.1 Summary of Objectives Achieved

This project produced and provided the client with five functional requirements. The five functional requirements included:

- System must include a budget analyst tool for counties and ZIP codes.
- System must have a geodatabase with sales and income tax data.
- System must display potential sales and income tax increase data results from budget analyst tool on individual choropleth maps for comparison.
- The user must be able to adjust percentages of sales and income tax increases.
- System must have a standard reporting template that produces maps.

The functional requirement to produce a standardized symbology for the budget analyst tool output feature class was not completed. This was because the symbology would have limited the user from viewing the data in its original state and would have delivered a pre-determined notion about the data. Therefore, the user would have the flexibility to view the data at their chosen discretion. The budget analyst tool was executed through the map document provided to the client and allowed the user to adjust percentages for sales tax increases. Using data from the geodatabase to run the tool, a derived income tax increase was derived and an output feature class was added to the map document. Also provided was a standard reporting template that produced maps for analysis.

7.2 Future Work

Future work associated with this project are discussed in the following sections as potential ideas to build upon the completed budget analyst tool.

7.2.1 Expansion of Budget Analyst Tool

The budget analyst tool used only income and sales tax data since it was readily available. The tool would be expanded to include the corporation tax, highway user taxes, motor vehicle fees, insurance tax, liquor tax, and tobacco taxes. These remaining taxes produce approximate 28% of revenue for the state and would complete the budget analyst tool if included. The equation provided by the client for the budget analyst tool would be modified to include all revenue sources and produce maps depicting the effects of the tax increases.

7.2.2 Budget Expenditure Tool

Expenditures in the state budget were not included in the budget analyst tool due to data and time constraints. Using expenditure data and providing a tool similar to the budget

analyst tool would help the client analyze the benefits provided by the state. Since there are multiple expenditures, providing maps that contain data of the people benefiting from them would help in understanding where the money is spent on county and ZIP code levels. This would be beneficial in determining any spatial relationships between family size, income groups, and resident locations based on those who benefit from the expenditures.

7.2.3 Customized Budget Analyst Tool Website

Once the revenue and expenditures tools are combined into a single tool, a customized website could be developed. The website would provide the user with comparison maps that have drop down lists containing the various revenue sources to be display on one map, and the second drop down list would contain the expenditures for display on the second map. The user would be able to adjust the selected revenue or expenditure percentage rates using a slide toolbar ranging from 1 to 100%. While the slide toolbar is adjusted, the comparison maps would change instantly for analysis. This would be helpful to those who use it since it provides results immediately and does not require any knowledge of any Esri software.

7.2.4 Addition of Sales Tax Industry Precision

Sales tax data used in this project did not define specific industries of where and what people were spending their money on. Using data that contains industry information in addition to the budget analyst tool, the user could create an output feature class of which industries generate more sales and what people are spending money on. An output example would show if people are spending money on shoes rather than on groceries. Another example would show if people are purchasing their taxable items at a mall versus at a big box retailer. Spatial statistics could be performed based on the findings.

7.3 Summary

The project concluded with solving the client's problem by providing a budget analyst tool, a geodatabase containing the data, and a standard reporting template. Future work recommended would be to improve the tool's capability by expanding the tool to include all expenditures and revenues, creating a customized website, and adding industry types to see on what and where people are spending their money.

Works Cited

- Berry, J. K. (1994). What can GIS do for you? *GIS World*, 24.
- Bolstad, P. (2008). *GIS fundamentals: A first text on geographic information systems*, (3rd ed.). White Bear Lake, Minnesota: Eider Press.
- California State Board of Equalization. (2011, November 1). *California City and County Sales and Use Tax Rates*. Retrieved from California State Board of Equalization: <http://www.boe.ca.gov/pdf/pub71.pdf>
- Cocklin, C., & Furuseth, O. (1994). Geographical dimensions of environmental restructuring in New Zealand. *The Professional Geographer*, 46(4), 459-467.
- Horn, M. (1999). GIS and the geography of politics. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems volume 2: management issues and applications* (2nd ed., Vol. II, pp. 939-951). New York: Wiley.
- Mitchell, A. (2005). *The ESRI guide to GIS analysis volume 2: spatial measurements & statistics* (Vol. II). Redlands: ESRI Press.
- O'Looney, J. (2000). *Beyond Maps: GIS and decision making in local government*. Redlands, California: ESRI Press.
- Office of Senate Counsel & Research for the Minnesota Senate. (2002). *1990 supreme court redistricting decisions*. Retrieved from Office of Senate Counsel & Research for the Minnesota Senate: <http://www.senate.leg.state.mn.us/departments/scr/redist/red907.htm#Illinois>
- Peters, A. H. & MacDonald, H. I. (2004). *Unlocking the census with GIS*. Redlands, California: ESRI Press.
- Reuters. (2011, February 4). *Analysis: Texas vs California: A tale of two budget deficits*. Retrieved from Reuters: <http://www.reuters.com>
- State of California. (2011). *2011-12 California State budget*. Retrieved from <http://www.ebudget.ca.gov/pdf/Enacted/BudgetSummary/FullBudgetSummary.pdf>
- Tomlinson, R. (2007). *Thinking about GIS: geographic information system planning for managers* (3rd ed.). Redlands, CA: ESRI Press.
- United States Census Bureau. (2008). *A compass for understanding and using American community survey data: What general data users need to know*. Washington D.C.: U.S. Government Printing Office.
- United States Census Bureau. (2011). *United States Census 2010*. Retrieved from <http://2010.census.gov>
- United States Department of Health & Human Services. (2011). *The personal responsibility and work opportunity reconciliation act of 1996*. Retrieved from <http://www.acf.hhs.gov/programs/cse/pubs/1996/news/prwora.htm>
- United States Government. (2011). *Recovery.gov - Track the money*. Retrieved from <http://www.recovery.gov>
- Wilson, D. (2009). Introduction: Racialized poverty in U.S. cities: Toward a refined racial economy perspective. *The Professional Geographer*, 61(2), 139-149.

Appendix A. County Budget Analyst Tool Python Script

```
import arcpy
from arcpy import env
env.workspace = "E:/MIP/CaliforniaBudgetCuts.gdb"

#Parameters using getParametersAsText()
CountySalesTax0910 = arcpy.GetParameterAsText(0)
StateSalesTaxRev = arcpy.GetParameterAsText(1)
CountyIncomeMadeandAGI = arcpy.GetParameterAsText(2)
Adjustedgrossincome = arcpy.GetParameterAsText(3)
CountyIncomeMade = arcpy.GetParameterAsText(4)
Percentmoneymade = arcpy.GetParameterAsText(5)
PercentMoneySpentSales = arcpy.GetParameterAsText(6)
#targetBudget = 72440557000 #72 Billion
targetBudget = int(arcpy.GetParameterAsText(7))
salesTaxPercent = float(arcpy.GetParameterAsText(8)) / 100.0

sumCountySalesTax0910 = 0.0

rows = arcpy.SearchCursor(CountySalesTax0910)
for row in rows:
    sumCountySalesTax0910 = sumCountySalesTax0910 +
    row.getValue(StateSalesTaxRev)
#arcpy.AddMessage("State Sales Tax Revenue = " + str(sumCountySalesTax0910))

targetTaxIncome = targetBudget - (sumCountySalesTax0910 * salesTaxPercent)
#arcpy.AddMessage("The taxable income target is " + str(targetTaxIncome))
del rows

percentMoneySpentRows = arcpy.SearchCursor(CountySalesTax0910)
sumPercentMoneySpent = 0
maxMoneySpent = -999999999999

for row in rows:
    sumPercentMoneySpent = sumPercentMoneySpent +
    row.getValue(PercentMoneySpentSales)
    if row.getValue(PercentMoneySpentSales) > maxMoneySpent:
        maxMoneySpent = row.getValue(PercentMoneySpentSales)
#arcpy.AddMessage("The percent money spent equals " + str(maxMoneySpent))

incomeRows = arcpy.SearchCursor(CountyIncomeMadeandAGI)
sumIncome = 0

for row in incomeRows:
```

```

sumIncome = sumIncome + row.getValue(Adjustedgrossincome)

#arcpy.AddMessage("The sum Income equals " + str(sumIncome))

percentMoneyMadeRows = arcpy.SearchCursor(CountyIncomeMade)
sumPercentMoneyMade = 0
maxPercentMoneyMade = -9999999999999999
for row in percentMoneyMadeRows:
    sumPercentMoneyMade = sumPercentMoneyMade +
row.getValue(Percentmoneymade)
    if row.getValue(Percentmoneymade) > maxPercentMoneyMade:
        maxPercentMoneyMade = row.getValue(Percentmoneymade)
#arcpy.AddMessage("The sum of percent money made equals " +
str(sumPercentMoneyMade))

incomeTaxPercent = (float(targetTaxIncome) / (float(sumIncome) +
float(sumPercentMoneyMade))) * 100
#incomeTaxPercent = (float(targetTaxIncome) / (float(sumIncome))) * 100
if incomeTaxPercent < 0:
    arcpy.AddError("Income tax percentage cannot be negative. Please insert another
percentage for analysis.")
else:
    arcpy.AddMessage("The income tax percentage should be " + str(incomeTaxPercent))

arcpy.SetParameterAsText(9,incomeTaxPercent)
arcpy.SetParameterAsText(10,maxMoneySpent)
arcpy.SetParameterAsText(11,maxPercentMoneyMade)

```

Appendix B. ZIP Code Budget Analyst Tool Python Script

```
import arcpy
from arcpy import env
env.workspace = "E:/MIP/CaliforniaBudgetCuts.gdb"

#Parameters using getParametersAsText()
zipCodeSalesRevenueMade = arcpy.GetParameterAsText(0)
StateSalesTaxRev = arcpy.GetParameterAsText(1)
zipCodeIncome_Made = arcpy.GetParameterAsText(2)
Adjustedgrossincome = arcpy.GetParameterAsText(3)
PercentageOfMoneyWeMake = arcpy.GetParameterAsText(4)
ZipCodeTotalTaxRevPercenIncome = arcpy.GetParameterAsText(5)
#targetBudget = 72440557000 #72 Billion
targetBudget = int(arcpy.GetParameterAsText(6))
salesTaxPercent = float(arcpy.GetParameterAsText(7)) / 100.0

sumZipCodeSalesRevenueMade = 0.0

rows = arcpy.SearchCursor(zipCodeSalesRevenueMade)
for row in rows:
    sumZipCodeSalesRevenueMade = sumZipCodeSalesRevenueMade +
    row.getValue(StateSalesTaxRev)
    arcpy.AddMessage("State Sales Tax Revenue = " + str(sumZipCodeSalesRevenueMade))

targetTaxIncome = targetBudget - (sumZipCodeSalesRevenueMade * salesTaxPercent)
arcpy.AddMessage("The taxable income target is " + str(targetTaxIncome))
del rows

percentMoneySpentRows = arcpy.SearchCursor(zipCodeSalesRevenueMade)
sumPercentMoneySpent = 0
maxMoneySpent = -999999999999

for row in rows:
    sumPercentMoneySpent = sumPercentMoneySpent +
    row.getValue(ZipCodeTotalTaxRevPercenIncome)
    if row.getValue(ZipCodeTotalTaxRevPercenIncome) > maxMoneySpent:
        maxMoneySpent = row.getValue(ZipCodeTotalTaxRevPercenIncome)
    arcpy.AddMessage("The percent money spent equals " + str(maxMoneySpent))

incomeRows = arcpy.SearchCursor(zipCodeIncome_Made)
sumzipCodeIncome_Made = 0

for row in incomeRows:
```

```

sumzipCodeIncome_Made = sumzipCodeIncome_Made +
row.getValue(Adjustedgrossincome)

arcpy.AddMessage("The sum Income equals " + str(sumzipCodeIncome_Made))

percentMoneyMadeRows = arcpy.SearchCursor(zipCodeIncome_Made)
sumPercentMoneyMade = 0
maxPercentMoneyMade = -9999999999999999
for row in percentMoneyMadeRows:
    sumPercentMoneyMade = sumPercentMoneyMade +
row.getValue(PercentageOfMoneyWeMake)
    if row.getValue(PercentageOfMoneyWeMake) > maxPercentMoneyMade:
        maxPercentMoneyMade = row.getValue(PercentageOfMoneyWeMake)
arcpy.AddMessage("The sum of percent money made equals " +
str(sumPercentMoneyMade))

incomeTaxPercent = (float(targetTaxIncome) / (float(sumzipCodeIncome_Made) +
float(sumPercentMoneyMade))) * 100
#incomeTaxPercent = (float(targetTaxIncome) / (float(sumIncome))) * 100
if incomeTaxPercent < 0:
    arcpy.AddError("Income tax percentage cannot be negative. Please insert another
percentage for analysis.")
else:
    arcpy.AddMessage("The income tax percentage should be " + str(incomeTaxPercent))

arcpy.SetParameterAsText(8,incomeTaxPercent)
arcpy.SetParameterAsText(9,maxMoneySpent)
arcpy.SetParameterAsText(10,maxPercentMoneyMade)

```