Visualize United States Personal Income and Expenditure data and Analyze the Possible Relationship

Haowei Sun

Contents

Summary ……..…… 3

Project Flowchart ……..…… 3

Datasets ……..…….. 4

- US States Personal Income ……..……... 4
- US State Expenditure ……..……... 4

Database Design ……..……..…..……... 5

- ER-Diagram ……..……... 5
- Relational Vocab ……..…….……... 6
- Sample Table ……..…….……... 6

Database Creation ……..…….……..... 7

Data Import ……..…….……..... 9

PCA ……..…….……..... 10

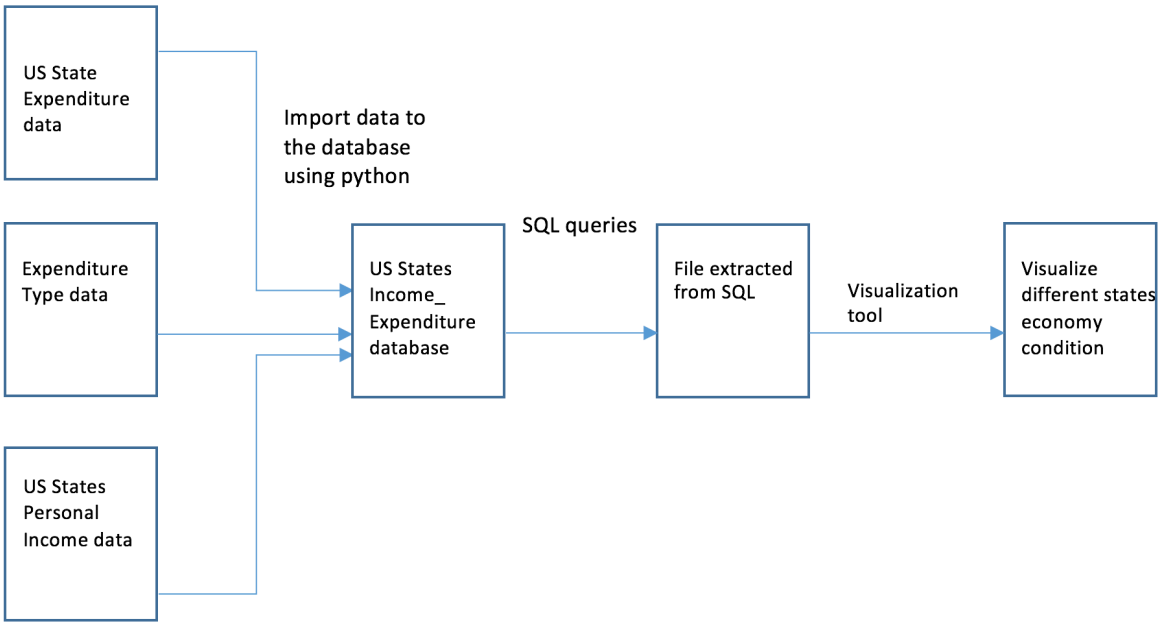
Visualization using Tableau ……..……..... 11

Challenges ……..…….……..... 15

Summary

The project demonstrates the personal income data and different kinds of expenditure data of all US states and analyses the relationship between them. There are several expenditure types and I use the principle component Analysis to select principal expenditure types and use map and graphs in Tableau to visualize the analysis. All the expenditure and personal income data can be visualized in the Tableau map and it is easy to compare the difference between different states using the map. For the specific state, I get the personal income and expenditure number in chart, suppose they have linear relationship and convince it in MATLAB.

Project Flowchart



The above figure represents the flowchart for our project workflow.

1. Data files: For this project we will be using data from multiple data sources. The US State Expenditure data is the file recording different states different kinds of expenditure in years, such as in Texas, goods expenditure in 2000.The Expenditure Type file contains 28 different types of Expenditure, such as goods, health. It is used to describe the attribute of the Expenditure. The US States Personal Income file is the quarter personal income of different states from 1948 to 2014. I sum the 4 quarters to the year data and choose the year from 1997 to 2012 in this project.
2. US States Income_ Expenditure Database: The data files are exported from csv into the database using python.
3. Files extracted from SQL queries: Based on the visualization task and the data required for it, make

appropriate SQL queries to the database and export the results in the form of a csv which can be later used for visualization.

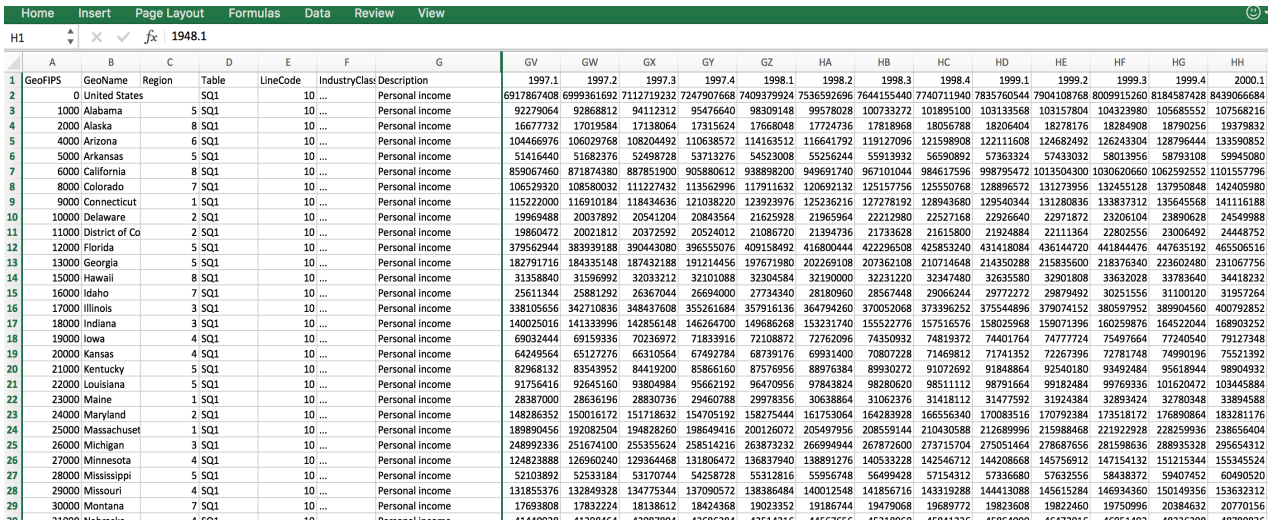
4. Analysis and Visualization: Using tools like MATLAB and Tableau to analyze the data and visualize them liking showing conditions of different states on the US map.

Datasets:

- US States Personal Income

Link: http://www.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=3#reqid=70&step=1&isuri=1

Bureau of Economic Analysis provides quarterly personal income data regionally from 1948 to 2014. I choose the data from 1997 to 2012 and for the same format as Expenditure, I sum the quarter income to the year income. I downloaded the excel from the website and part of the excel is below. The meta data unit is thousand.



Home Insert Page Layout Formulas Data Review View

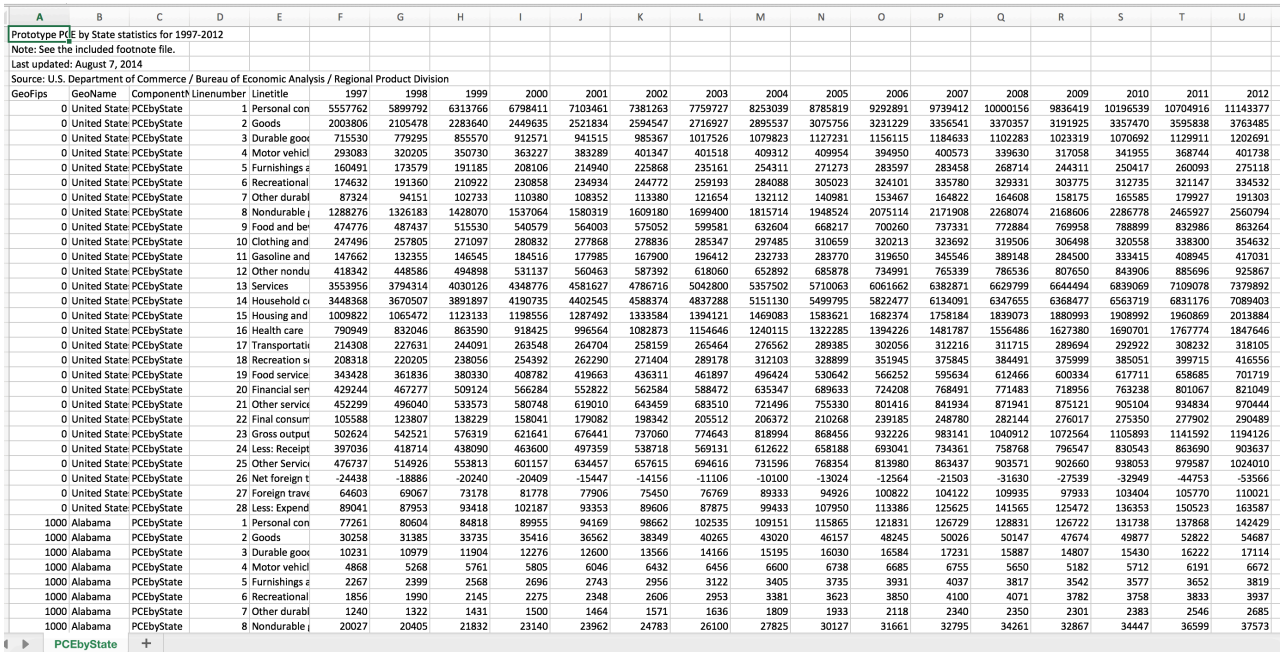
H1 fx 1948.1

	A	B	C	D	E	F	G	GV	GW	GX	GY	GZ	HA	HB	HC	HD	HE	HF	HG	HH
1	GeoFIPS	GeoName	Region	Table	LineCode	IndustryClass	Description	1997.1	1997.2	1997.3	1997.4	1998.1	1998.2	1998.3	1998.4	1999.1	1999.2	1999.3	1999.4	2000.1
2	0	United States		SQ1	10	...	Personal income	6917867408	6999361692	7112719232	7247907668	7409379924	7536592696	7644155440	7740711940	7835760544	7904108768	8009915260	8184587428	8439066684
3	1000	Alabama	5	SQ1	10	...	Personal income	92279064	92868812	94112312	95476640	98309148	99578028	100733272	101895100	103133568	103157804	104323980	105685552	107568216
4	2000	Alaska	8	SQ1	10	...	Personal income	16677732	17019584	17138064	17315624	17668048	17724736	17818968	18056788	18206404	18278176	18284908	18790256	19379832
5	4000	Arizona	6	SQ1	10	...	Personal income	104466976	106029768	108204492	110638572	114163512	116641792	119127096	121598908	122111608	124682492	126243304	128796444	133590852
6	5000	Arkansas	5	SQ1	10	...	Personal income	51416440	51682376	52498728	53713276	54523008	55256244	55913932	56590892	57363324	57433032	58013956	58793108	59945080
7	6000	California	8	SQ1	10	...	Personal income	859067460	871874380	887851900	905880612	938898200	949691740	967101044	984617596	998795472	1013504300	1030620660	1062592552	1101557796
8	8000	Colorado	7	SQ1	10	...	Personal income	106529320	108580032	111227432	113562996	117911632	120692132	125157756	125550768	128896572	131273956	132455128	137950848	142405980
9	9000	Connecticut	1	SQ1	10	...	Personal income	115222000	116910184	118434636	121038220	123923976	125236216	127278192	128943680	129540344	131280836	133837312	135645568	141116188
10	10000	Delaware	2	SQ1	10	...	Personal income	19969488	20037892	20541204	20843564	21625928	21965964	22212980	22527168	22926640	22971872	23206104	23890628	24549988
11	11000	District of Co	2	SQ1	10	...	Personal income	19860472	20021812	20372592	20524012	21086720	21394736	21733628	21615800	21924884	22111364	22802556	23006492	24448752
12	12000	Florida	5	SQ1	10	...	Personal income	379562944	383939188	390443080	396555076	409158492	416800444	422296508	425853240	431418084	436144720	441844476	447635192	465506516
13	13000	Georgia	5	SQ1	10	...	Personal income	182791716	184335148	187432188	191214456	197671980	202269108	207362108	210714648	214350288	215835600	218376340	223602480	231067756
14	15000	Hawaii	8	SQ1	10	...	Personal income	31358840	31596992	32033212	32101088	32304584	32190000	32231220	32347480	32635580	32901808	33632028	33783640	34418232
15	16000	Idaho	7	SQ1	10	...	Personal income	25611344	25881292	26367044	26694000	27734340	28180960	28567448	29066244	29772272	29879492	30251556	31100120	31957264
16	17000	Illinois	3	SQ1	10	...	Personal income	338105656	342710836	348437608	355261684	357916136	364794260	370052068	373396252	375544896	379074152	380597952	389904560	400792852
17	18000	Indiana	3	SQ1	10	...	Personal income	140025016	141333996	142856148	146264700	149686268	153231740	155522776	157516576	158025968	159071396	160259876	164522044	168903252
18	19000	Iowa	4	SQ1	10	...	Personal income	69032444	69159336	70236972	71833916	72108872	72762096	74350932	74819372	74401764	74777724	75497664	77240540	79127348
19	20000	Kansas	4	SQ1	10	...	Personal income	64249564	65127276	66310564	67492784	68739176	69931400	70807228	71469812	71741352	72267396	72781748	74990196	75521392
20	21000	Kentucky	5	SQ1	10	...	Personal income	82968132	83543952	84419200	85866160	87576956	88976384	89930272	91072692	91848864	92540180	93492484	95618944	98904932
21	22000	Louisiana	5	SQ1	10	...	Personal income	91756416	92645160	93804984	95662192	96470956	97843824	98280620	98511112	98791664	99182484	99769336	101620472	103445884
22	23000	Maine	1	SQ1	10	...	Personal income	28387000	28636196	28830736	29460788	29978356	30638864	31062376	31418112	31477592	31924384	32893424	32780348	33894588
23	24000	Maryland	2	SQ1	10	...	Personal income	148286352	150016172	151718632	154705192	158275444	161753064	164283928	166556340	170083516	170792384	173518172	176890864	183281176
24	25000	Massachuset	1	SQ1	10	...	Personal income	189890456	192082504	194828260	198649416	200126072	205497956	208559144	210430588	212689996	215988468	221922928	228259936	238656404
25	26000	Michigan	3	SQ1	10	...	Personal income	248992336	251674100	255355624	258514216	263873232	266994944	267872600	273715704	275051464	278687656	281598636	288935328	295654312
26	27000	Minnesota	4	SQ1	10	...	Personal income	124823888	126960240	129364468	131806472	136837940	138891276	140533228	142546712	144208668	145756912	147154132	151215344	155345524
27	28000	Mississippi	5	SQ1	10	...	Personal income	52103892	52533184	53170744	54258728	55312816	55956748	56499428	57154312	57336680	57632556	58438372	59407452	60490520
28	29000	Missouri	4	SQ1	10	...	Personal income	131855376	132849328	134775344	137090572	138386484	140012548	141856716	143319288	144413088	145615284	146934360	150149356	153632312
29	30000	Montana	7	SQ1	10	...	Personal income	17693808	17832224	18138612	18424368	19023352	19186744	19479068	19689772	19823608	19822460	19750996	20384632	20770156

- US State Expenditure

Link: http://www.bea.gov/newsreleases/regional/pce/pce_newsrelease.htm

Bureau of Economic Analysis provides 28 kinds of different expenditure for 51 states from 1997 to 2012. The data contains the states' GeoFips , states' name, type name of expenditure and the expenditure data. . I downloaded the excel from the website. The meta data unit is million.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Prototype PCE by State statistics for 1997-2012																				
2	Note: See the included footnote file.																				
3	Last updated: August 7, 2014																				
4	Source: U.S. Department of Commerce / Bureau of Economic Analysis / Regional Product Division																				
5	GeoFips	GeoName	ComponentN	Linenumber	Linetitle	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
6	0	United State	PCEbyState	1	Personal con	5557762	5899792	6313766	6798411	7103461	7381263	7759727	8253039	8785819	9292891	9739412	10000156	9836419	10196539	10704916	11143377
7	0	United State	PCEbyState	2	Goods	2003806	2105478	2283640	2449635	2521834	2594547	2716927	2895537	3075756	3231229	3356541	3370357	3191925	3357470	3595838	3763485
8	0	United State	PCEbyState	3	Durable gooc	715530	779295	855570	912571	941515	985367	1017526	1079823	1127231	1156115	1184633	1102283	1023319	1070692	1129911	1202691
9	0	United State	PCEbyState	4	Motor vehicl	293083	320205	350730	363227	383289	401347	401518	409312	409954	394950	400573	339630	317058	341955	368744	401738
10	0	United State	PCEbyState	5	Furnishings a	160491	173579	191185	208106	214940	225868	235161	254311	271273	283597	283458	268714	244311	250417	260093	275118
11	0	United State	PCEbyState	6	Recreational	174632	191360	210922	230858	234934	244772	259193	284088	305023	324101	335780	329331	303775	312735	321147	334532
12	0	United State	PCEbyState	7	Other durabl	87324	94151	102733	110380	108352	113380	121654	132112	140981	153467	164822	164608	158175	165585	179927	191303
13	0	United State	PCEbyState	8	Nondurable g	1288276	1326183	1428070	1537064	1580319	1609180	1699400	1815714	1948524	2075114	2171908	2268074	2168606	2286778	2465927	2560794
14	0	United State	PCEbyState	9	Food and bev	474776	487437	515530	540579	564003	575052	599581	632604	668217	700260	737331	772884	769958	788899	832986	863264
15	0	United State	PCEbyState	10	Clothing and	247496	257805	271097	280832	277868	278836	285347	297485	310659	320213	323692	319506	306498	320558	338300	354632
16	0	United State	PCEbyState	11	Gasoline and	147662	132355	146545	184516	177985	167900	196412	232733	283770	319650	345546	389148	284500	333415	408945	417031
17	0	United State	PCEbyState	12	Other nondu	418342	448586	494898	531137	560463	587392	618060	652892	685878	734991	765339	786536	807650	843906	885696	925867
18	0	United State	PCEbyState	13	Services	3553956	3794314	4030126	4348776	4581627	4786716	5042800	5357502	5710063	6061662	6382871	6629799	6644494	6839069	7109078	7379892
19	0	United State	PCEbyState	14	Household c	3448368	3670507	3891897	4190735	4402545	4588374	4837288	5151130	5499795	5822477	6134091	6347655	6368477	6563719	6831176	7089403
20	0	United State	PCEbyState	15	Housing and	1009822	1065472	1123133	1198556	1287492	1333584	1394121	1469083	1583621	1682374	1758184	1839073	1880993	1908992	1960869	2013884
21	0	United State	PCEbyState	16	Health care	790949	832046	863590	918425	996564	1082873	1154646	1240115	1322285	1394226	1481787	1556486	1627380	1690701	1767774	1847646
22	0	United State	PCEbyState	17	Transportati	214308	227631	244091	263548	264704	258159	265464	276562	289385	302056	312216	311715	289694	292922	308232	318105
23	0	United State	PCEbyState	18	Recreation s	208318	220205	238056	254392	262290	271404	289178	312103	328899	351945	375845	384491	375999	385051	399715	416556
24	0	United State	PCEbyState	19	Food service	343428	361836	380330	408782	419663	436311	461897	496424	530642	566252	595634	612466	600334	617711	658685	701719
25	0	United State	PCEbyState	20	Financial ser	429244	467277	509124	566284	552822	562584	588472	635347	689633	724208	768491	771483	718956	763238	801067	821049
26	0	United State	PCEbyState	21	Other service	452299	496040	533573	580748	619010	643459	683510	721496	755330	801416	841934	871941	875121	905104	934834	970444
27	0	United State	PCEbyState	22	Final consum	105588	123807	138229	158041	179082	198342	205512	206372	210268	239185	248780	282144	276017	275350	277902	290489
28	0	United State	PCEbyState	23	Gross output	502624	542521	576319	621641	676441	737060	774643	818994	868456	932226	983141	1040912	1072564	1105893	1141592	1194126
29	0	United State	PCEbyState	24	Less: Receipt	397036	418714	438090	463600	497359	538718	569131	612622	658188	693041	734361	758768	796547	830543	863690	903637
30	0	United State	PCEbyState	25	Other Servic	476737	514926	553813	601157	634457	657615	694616	731596	768354	813980	863437	903571	902660	938053	979587	1024010
31	0	United State	PCEbyState	26	Net foreign t	-24438	-18886	-20240	-20409	-15447	-14156	-11106	-10100	-13024	-12564	-21503	-31630	-27539	-32949	-44753	-53566
32	0	United State	PCEbyState	27	Foreign trave	64603	69067	73178	81778	77906	75450	76769	89333	94926	100822	104122	109935	97933	103404	105770	110021
33	0	United State	PCEbyState	28	Less: Expend	89041	87953	93418	102187	93353	89606	87875	99433	107950	113386	125625	141565	125472	136353	150523	163587
34	1000	Alabama	PCEbyState	1	Personal con	77261	80604	84818	89955	94169	98662	102535	109151	115865	121831	126729	128831	126722	131738	137868	142429
35	1000	Alabama	PCEbyState	2	Goods	30258	31385	33735	35416	36562	38349	40265	43020	46157	48245	50026	50147	47674	49877	52822	54687
36	1000	Alabama	PCEbyState	3	Durable gooc	10231	10979	11904	12276	12600	13566	14166	15195	16030	16584	17231	15887	14807	15430	16222	17114
37	1000	Alabama	PCEbyState	4	Motor vehicl	4868	5268	5761	5805	6046	6432	6456	6600	6738	6685	6755	5650	5182	5712	6191	6672
38	1000	Alabama	PCEbyState	5	Furnishings a	2267	2399	2568	2696	2743	2956	3122	3405	3735	3931	4037	3817	3542	3577	3652	3819
39	1000	Alabama	PCEbyState	6	Recreational	1856	1990	2145	2275	2348	2606	2953	3381	3623	3850	4100	4071	3782	3758	3833	3937
40	1000	Alabama	PCEbyState	7	Other durabl	1240	1322	1431	1500	1464	1571	1636	1809	1933	2118	2340	2350	2301	2383	2546	2685
41	1000	Alabama	PCEbyState	8	Nondurable g	20027	20405	21832	23140	23962	24783	26100	27825	30127	31661	32795	34261	32867	34447	36599	37573

PCEbyState +

States: 51 states in U.S., labeled by its own GeoFips

Year: 1997 to 2012

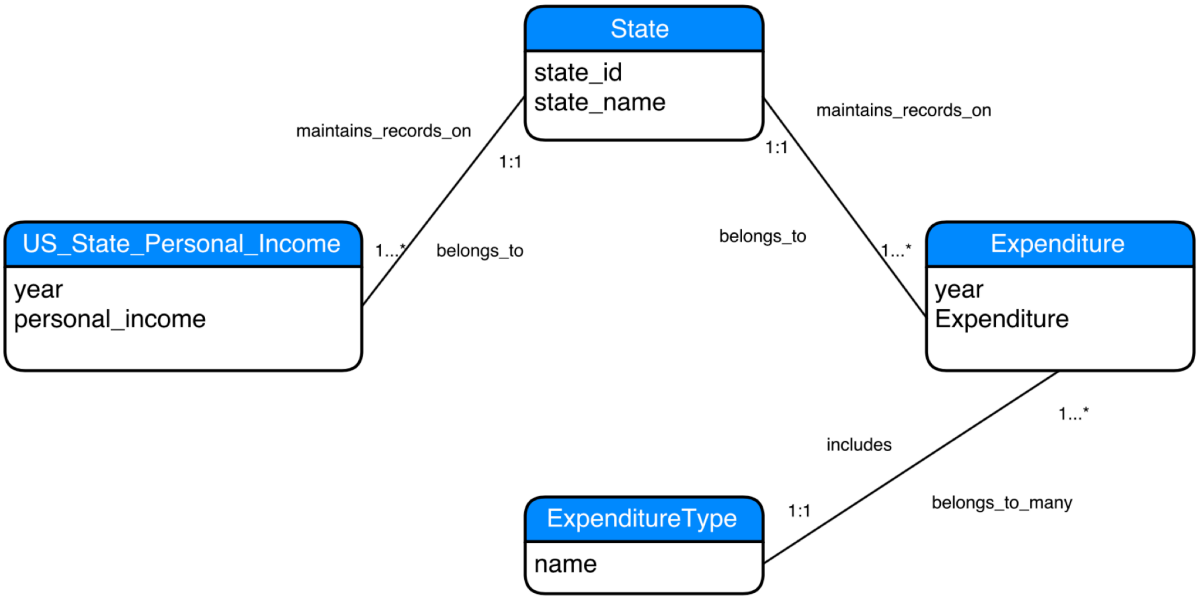
Database Design

- ER-Diagram

The ER-Diagram is built to connect the 4 tables together and find the connection between them. The design tries to find the affects between the personal income and the expenditure in different U.S. states. For the personal income table, personal income of the is used to measure the living condition of people in a specific state and each row in the personal income table correspond to a specific year for any giving state. This results in a ‘has and belongs’ to many relationship between US States Personal Income and State tables with state id as a foreign key in US States Personal Income table.

For the Expenditure table, each expenditure is associated with a state and a state can have multiple expenditure records, so I have a ‘has and belongs’ to many relationship between the State and Expenditure table using state id as foreign key. Also, for the expenditure, there are different types to describe the data and each row distinctly refers only one type Expenditure Type. I build an Expenditure Type table and there is a ‘has and belongs’ to many relationship between Expenditure and Expenditure Type using Expenditure Type id as the foreign key. Since every state needs to statistic its different kinds of expenditure for the calculation of state GDP, all states have 28 kinds of expenditure every year.

The ER-diagram, relational vocab and table sketches are provided in the following sections:



- Relational Vocab

State: has_many US_State_Personal_Income

US_State_Personal_Income: belongs_to State

State: has_many_ Expenditure

Expenditure: belongs_to State

ExpenditureType: has_many_ Expenditure

Expenditure: belongs_to ExpenditureType

- Sample Table

State

state_id	state_name
1000	Alabama
2000	Alaska

4000	Arizona

US_State_Personal_Income(million)

id	state_id	year	personal_income
1	1000	1997	374736
2	1000	1998	400515
3	1000	1999	416300

Expenditure(million)

id	state_id	year	expenditureType_id	expenditure
1	1000	1997	1	77261
2	1000	1998	1	80604
3	1000	1999	1	84818

ExpenditureType

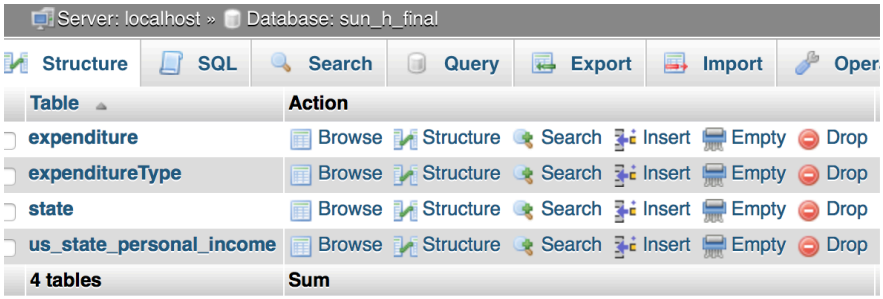
expenditureId	name
1	Personal consumption expenditures
2	Goods
3	Durable goods

Us_state_personal_income_expenditure(million)

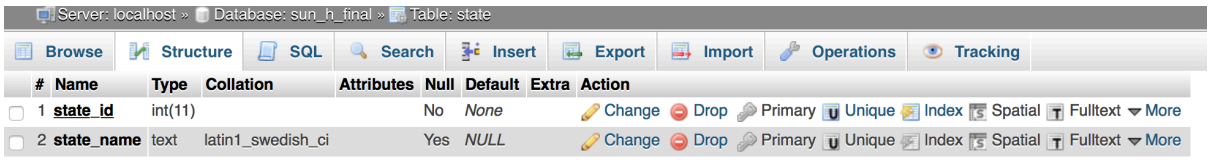
id	state_id	state_name	year	personal_income	expenditureType	expenditure
1	1000	Alabama	1997	374736	Personal consumption expenditures	77261
2	1000	Alabama	1998	400515	Personal consumption expenditures	80604
3	1000	Alabama	1999	416300	Personal consumption expenditures	84818

Database creation

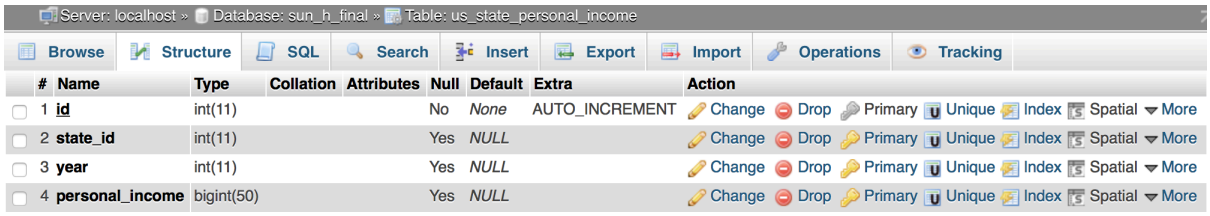
The database was created using PhpMyAdmin.



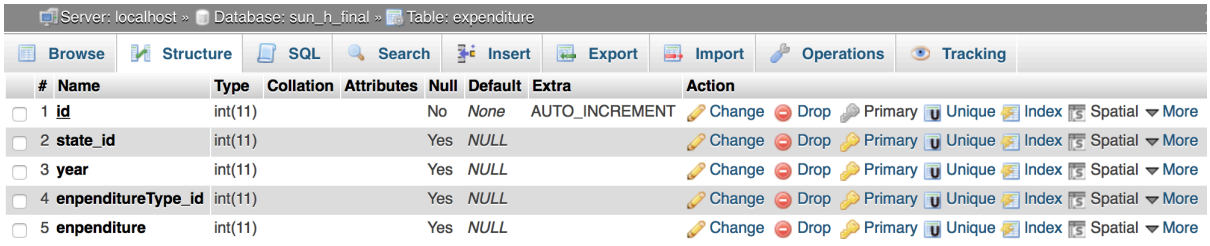
State



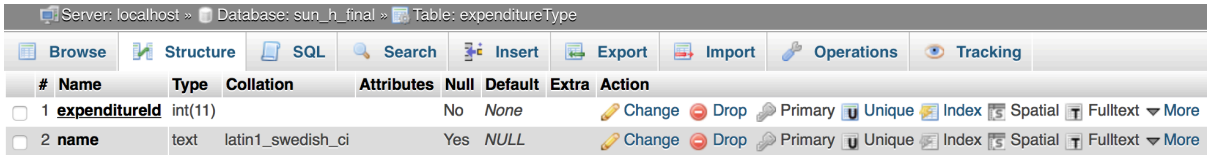
US_State_Personal_Income



Expenditure



ExpenditureType



Data Import

- Read and Insert csv data to the database

Connect to the server:

```
import pymysql.cursors
import pprint
import csv

connection = pymysql.connect(host="localhost",
                user="sun_h",
                 passwd="82562533s",
                 db="sun_h_final",
                 autocommit=True,
                 cursorclass=pymysql.cursors.DictCursor)

#connect to the server and read the first csv file
cursor = connection.cursor()
with open('SQ1_1948_2015.csv') as csvfile:
   myCSVReader = csv.DictReader(csvfile, delimiter=",", quotechar="")
```

read the 1st csv and insert the choosing data to the database:

```
   #for the state table, choose the state GeoFips as the state_id and the primary key and put the state_id and state_name to state table
   for row in myCSVReader:
      if  row["GeoFIPS"].isdigit() and int(row["GeoFIPS"]) >= 1000 and int(row["GeoFIPS"]) <=56000:
         state_sql = "INSERT INTO state( state_name, state_id) VALUE (%(dict_name)s, %(dict_stateid)s)"
         param_dict = { "dict_name":row["GeoName"], "dict_stateid":row["GeoFIPS"]}
         cursor.execute(state_sql, param_dict)


  #for the table us_state_personal_income, choose the year  from 1997 to 2013 increasing 1 per time
  #the statistics about the personal income is for the season and count the four quarters of the income for the whole year
  #input the state_id, year and the personal income of states to the table
  for year in range(1997, 2013, 1):
     income = 0
     for session in {0.1, 0.2, 0.3, 0.4} :
        if  year + session > 1997 and year + session < 2013:
           income += int(row[str(year + session)])
     income_sql = "INSERT INTO us_state_personal_income(state_id, year, personal_income) VALUE (%(dict_stateid)s, %(dict_year)s, %(dict_income)s)"
     param_dict = {"dict_stateid":row["GeoFIPS"], "dict_year":year, "dict_income":income}
     cursor.execute(income_sql, param_dict)
```

read the 2nd csv and build another two database table:

```
#read the PCEState csv and skip the first 4 lines since they are the basic information of the csv and not useful for the database
dict_expense = list()
with open('PCEbyState.csv') as csvfile:
   csvReader = csv.reader(csvfile, delimiter = ",", quotechar = "|")
   for i in range(4) :
      line = next(csvReader)


   #for the expenditure type table, there are 28 kinds of expenditure  and  I put the type in the expenditure table
   myCSVReader = csv.DictReader(csvfile, delimiter=",", quotechar="")
   for row in myCSVReader:
      if  row["GeoFips"].isdigit() and int(row["GeoFips"]) >= 1000 and int(row["GeoFips"]) <=56000:
         if row["Linetitle"] not in dict_expense:
            dict_expense.append(row["Linetitle"])
            income_sql = "INSERT INTO expenditureType(expenditureId, name) VALUE (%(dict_expenseId)s, %(dict_expense)s)"
            param_dict = {"dict_expenseId":row["Linenumber"], "dict_expense":row["Linetitle"]}
            cursor.execute(income_sql, param_dict)
```

```
#for the expenditure table, the data is large since each record is differnt in state_id, expenditure type and year and the year is from 1997 to 2012
for year in range(1997, 2013, 1):
    state_sql2 = "INSERT INTO expenditure(state_id, year, enpenditure, enpenditureType_id) VALUE (%(dict_stateid)s, %(dict_year)s, %(dict_expenditure)s, %(dict_expenditure_type
    param_dict2 = {"dict_stateid":row["GeoFips"], "dict_year": year, "dict_expenditure": row[str(year)], "dict_expenditure_type": int(row["Linenumber"])}
    cursor.execute(state_sql2, param_dict2)
```

PCA

Since there are 28 types expenditure, I want to choose the principal effective elements for the expenditure and I use the principle component, since it is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables.

I choose to do PCA state by state in MATLAB, just using the "insert data" in MATLAB inserting the excel file to MATLAB and clicking the "numeric data" to delete the title.

Variables – expenditure

expenditure

22848x5 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1000	1997	1	77261							
2	2	1000	1998	1	80604							
3	3	1000	1999	1	84818							
4	4	1000	2000	1	89955							
5	5	1000	2001	1	94169							
6	6	1000	2002	1	98662							
7	7	1000	2003	1	102535							
8	8	1000	2004	1	109151							
9	9	1000	2005	1	115865							
10	10	1000	2006	1	121831							
11	11	1000	2007	1	126729							
12	12	1000	2008	1	128831							
13	13	1000	2009	1	126722							

After importing the expenditure to MATLAB, for the first state in the data:

```
>> m = reshape(expenditure(1:448,5),16,28);
>> [COEFF,SCORE] = princomp(m);
```

The matrix is below:

Variables – m

expenditure | m

16x28 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	77261	30258	10231	4868	2267	1856	1240	20027	7743	3552	2438	629
2	80604	31385	10979	5268	2399	1990	1322	20405	7855	3696	2121	673
3	84818	33735	11904	5761	2568	2145	1431	21832	8267	3749	2341	747
4	89955	35416	12276	5805	2696	2275	1500	23140	8320	3807	3049	796
5	94169	36562	12600	6046	2743	2348	1464	23962	8781	3798	2853	853
6	98662	38349	13566	6432	2956	2606	1571	24783	9038	3897	2667	918
7	102535	40265	14166	6456	3122	2953	1636	26100	9328	3824	3149	979
8	109151	43020	15195	6600	3405	3381	1809	27825	9672	3995	3728	1043
9	115865	46157	16030	6738	3735	3623	1933	30127	10254	4214	4557	1110
10	121831	48245	16584	6685	3931	3850	2118	31661	10260	4308	5284	1180
11	126729	50026	17231	6755	4037	4100	2340	32795	10565	4338	5633	1225
12	128831	50147	15887	5650	3817	4071	2350	34261	11017	4326	6267	1265
13	126722	47674	14807	5182	3542	3782	2301	32867	11004	4221	4514	1312
14	131738	49877	15430	5712	3577	3758	2383	34447	11127	4397	5320	1360
15	137868	52822	16222	6191	3652	3833	2546	36599	11610	4564	6433	1399
16	142429	54687	17114	6672	3819	3937	2685	37573	11792	4625	6424	1473
17												

I get the COEFF, which returns the principal component coefficients and SCORE, which returns the principal component scores, the representation of X in the principal component space.

```
>> [COEFF,SCORE,latent,tsquare] = princomp(m);
>> cumsum(latent)./sum(latent);
```

I calculate the latent, which returns the principal component variances, i.e., the eigenvalues of the covariance matrix m.

```
ans =

    0.9964
    0.9988
    0.9994
    0.9997
    0.9999
    0.9999
    1.0000
```

The answer means that the first principal component can express 99.64% of the 28 types and the first seven principal components can express almost 100% of the 28 types expenditure. It means the first principal component can be used to analyze the relationship between the expenditure and personal income.

Visualization using Tableau

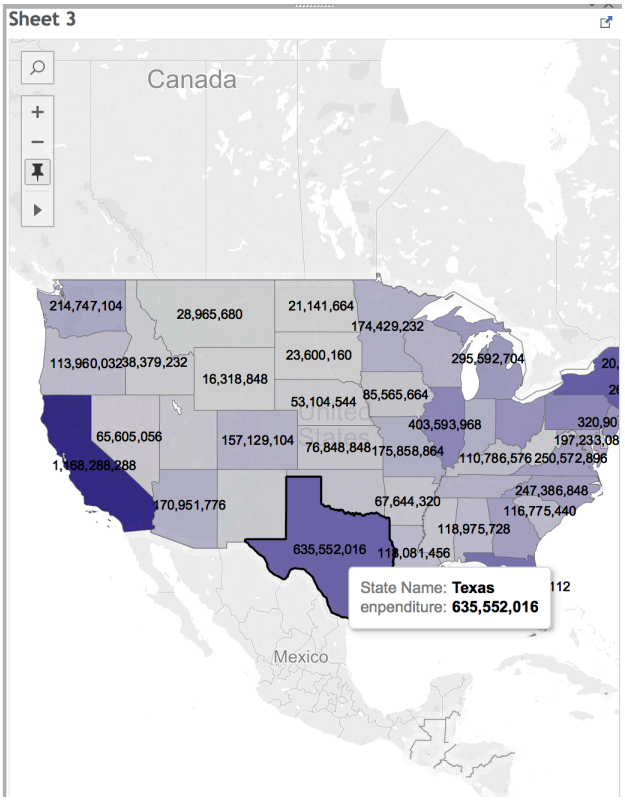
To visualize the results for our analysis, I decided that it would be visually appealing if I could look at the results on a map view. In order to achieve that, I found in sample report in Canvas the author using Tableau as the best fit for visualization. Tableau can easily load data from Microsoft excel files, text file, csv files and even database connections via Microsoft Access. The results from my queries were exported from the database using Python into a csv file. I directly used this file to feed in as an input to Tableau. Tableau is a very powerful tool and is capable of providing different forms of visualization from line graphs, bar graphs, pie charts, to gantt charts, treemaps and box-and-whisker plots. The type of graph that was most relevant to us was the symbol maps or filled maps. These maps require a geographic dimension in order to be populated for the view. Another interesting feature that we found in this was its ability to identify regions on the map.

Inserting all the expenditure data to Tableau, it is easy to find the different states expenditure condition and since it uses the color extent to show the amount compared to other states. For the direct watching, it shows that states having a larger

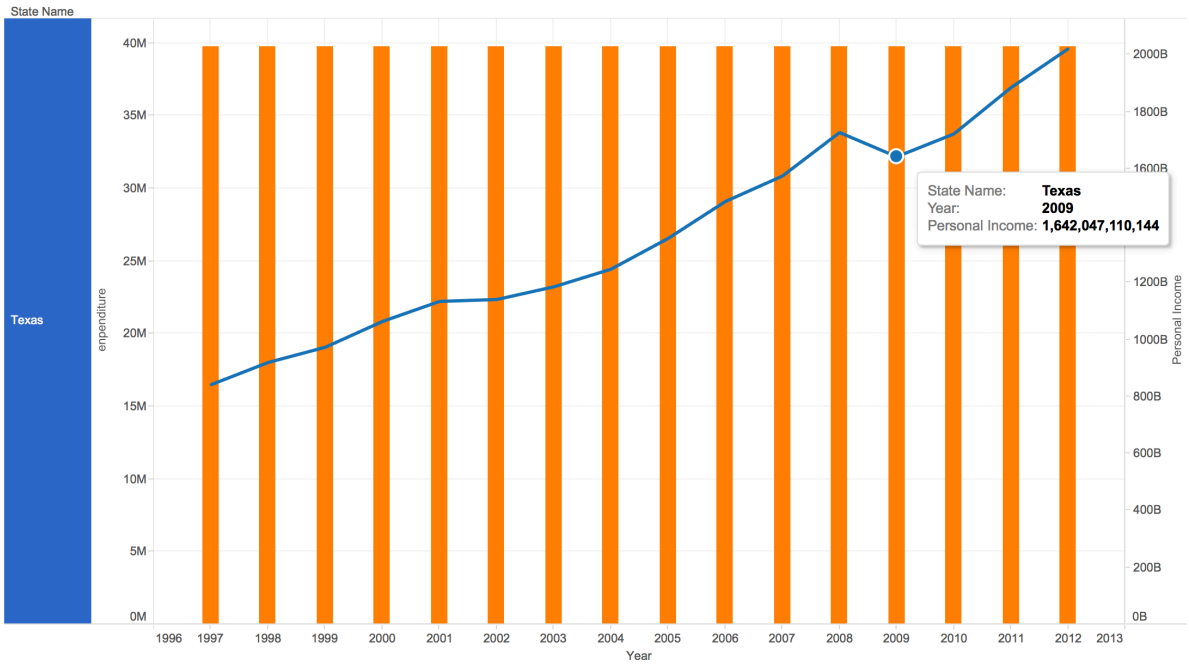
total expenditure most locating in the border of the US like California, Texas, New York and Florida. States locating in the middle of U.S., usually have a lighter color, which means the expenditure of states is less. When clicking specific state in the map, I can get the name and the total expenditure number of the state.

Meanwhile, on the left of the map, all different expenditure type numbers are shown and I can get every type specific type number.

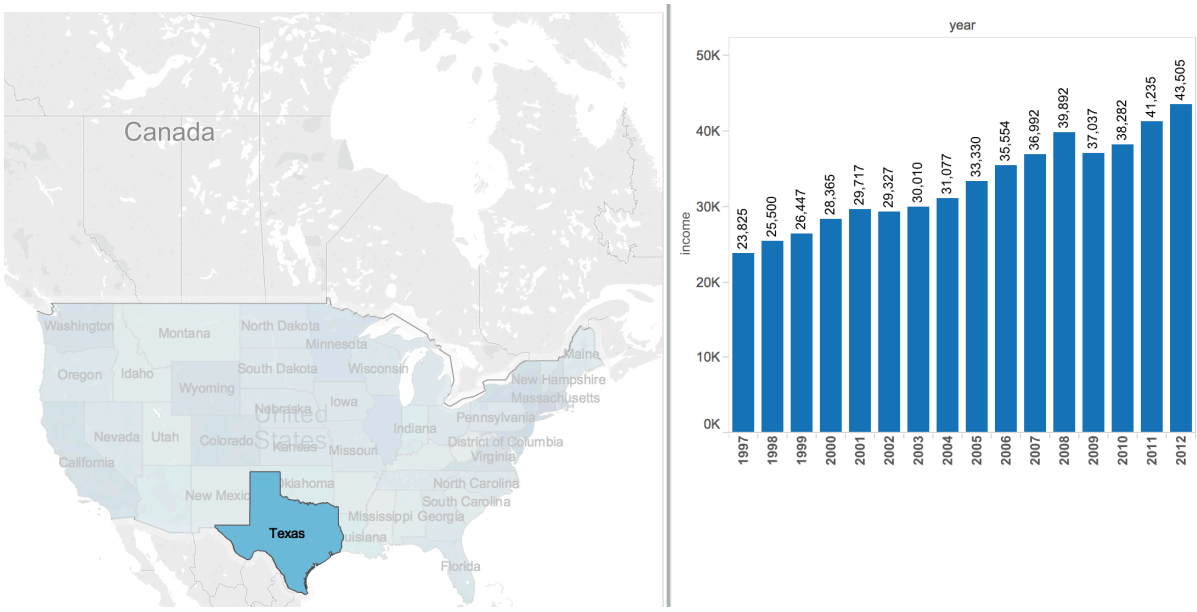
Expenditure Id	Name	
1	Personal consumption expenditures	2,156,267,984
2	Goods	744,223,904
3	Durable goods	260,544,944
4	Motor vehicles and parts	94,357,040
5	Furnishings and durable household equipment	60,809,872
6	Recreational goods and vehicles	70,355,056
7	Other durable goods	35,023,328
8	Nondurable goods	483,678,720
9	Food and beverages purchased for off-premises co..	168,373,760
10	Clothing and footwear	76,653,344
11	Gasoline and other energy goods	66,689,824
12	Other nondurable goods	171,962,016
13	Services	1,412,043,952
14	Household consumption expenditures (for services)	1,357,402,176
15	Housing and utilities	392,148,112
16	Health care	329,079,920
17	Transportation services	71,020,640
18	Recreation services	81,190,912
19	Food services and accommodations	131,073,808
20	Financial services and insurance	165,908,464
21	Other services	186,980,112
22	Final consumption expenditures of nonprofit instit..	54,641,568
23	Gross output of nonprofit institutions	217,426,192
24	Less: Receipts from sales of goods and services by nonpro..	162,784,656
25	Other Services, excluding net foreign travel	192,937,120
26	Net foreign travel	-5,956,832



Also, for every state, I can check each year personal income and expenditure changes. For example, for Texas, in the chart, the lines show the changes of the personal income over the years and the bars shows the expenditure changes over the year and it is easy to find the changing trend in the chart.



All the above graphs are the data for the whole states. However, for most people, they are more concerned about the average income they get per year. So the following graph reflect the condition of per capita personal income(person income divide by state population). When clicking the state name in the map, I can get the 1997 头



I suppose that there maybe a linear relationship between every state personal income and expenditure and I tries to select data using SQL use MATLAB to convince this relationship.

In SQL operation, I choose the state id, the year, the personal income(since it already has one value for one year), and I sum all types expenditure group by state and year in order to get the value of one total expenditure for one state per year.

In this way, I can get the value of personal income and expenditure of one state per year.

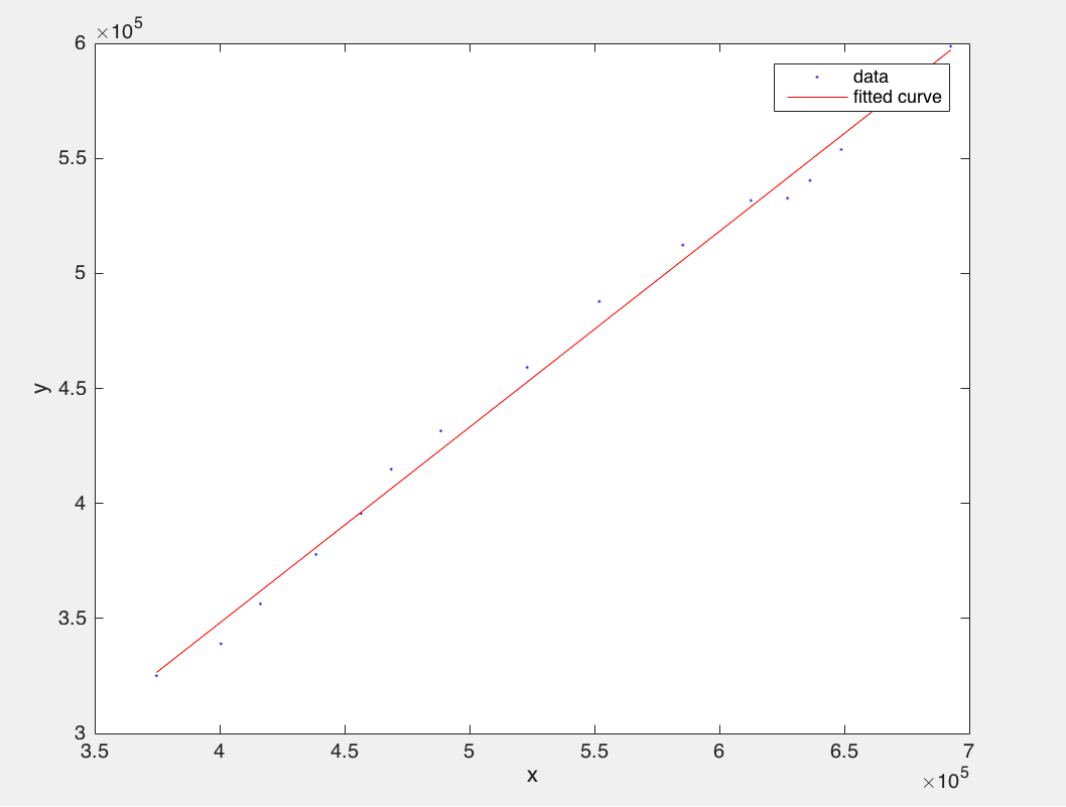
```
SELECT state.state_id, expenditure.year, us_state_personal_income.personal_income, SUM( expenditure.enpenditure ) AS expenditure
FROM state, us_state_personal_income, expenditure
WHERE expenditure.year = us_state_personal_income.year
AND state.state_id = us_state_personal_income.state_id
AND state.state_id = expenditure.state_id
GROUP BY expenditure.state_id, expenditure.year
ORDER BY state.state_id, expenditure.year
```

For example, for the first state, I select the personal income and expenditure and suppose personal income as x and expenditure as y.

```
>> p = fittype('poly1');
>> f = fit(x, y,p)

f =

     Linear model Poly1:
     f(x) = p1*x + p2
     Coefficients (with 95% confidence bounds):
       p1 =      0.8513  (0.8149, 0.8876)
       p2 =        7488  (-1.239e+04, 2.737e+04)
>> poly(f,x,y);
```



From the plot, it shows an explicit linear relationship between the personal income and expenditure and when personal income increases, the expenditure will increase as well.

Challenges

When first doing this project, I found it really difficult to find the data and build the relationship since I hope the data can be as large as possible. Then I think the data related to the U.S. economy is usually listed according to the year and the state. So I choose the income and expenditure and suppose there is maybe a relationship between them and using the state id as the foreign key.

Meanwhile, at first, I didn't notice the different the different unit between the personal income and the expenditure. So there is a huge difference between the number level of these two table. Then I found for the personal income, the metadata unit is thousand, but for the expenditure, the unit is million. I think it is really convenient to change the personal income unit to million before inserting it to the database. So I made the change and the linear relationship between two of them is much more normal compared to the different units.

When importing the data into the database, the time level of the personal income and expenditure is different and I choose to add the quarters to the year and makes them the same. There are also many choices about how to display the data and I tried many ways, like PCA in the project. However, since PCA is the way to decrease the variable number and it can not be used to discuss the relationship between the personal income and expenditure, I use the original data in the Tableau.