Project Supplemental Information

Christian Campbell

2023-08-13

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.3.1
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
# First Data Set
## Setting the working directory to the root of my DSC 520 directory
setwd("C:/Users/Owner/Documents/Important/Bellevue/Statistics for Data Science/Project Data")

## Load the Household Income data set
median_income <- read.csv("Median_Household_Income.csv")
# Raw Median Income data set
head(median_income)
```

```
##            X Median.Income.by.State..2018.adjusted.dollars.     X.1     X.2
## 1                                                                1984    1985    1986
## 2    Alabama                                                    39,958  40,915  41,963
## 3     Alaska                                                    74,689  77,625  68,775
## 4    Arizona                                                    49,457  53,287  55,931
## 5   Arkansas                                                    36,181  38,946  41,082
## 6 California                                                    58,372  60,215  63,629
##       X.3     X.4     X.5     X.6     X.7     X.8     X.9    X.10    X.11
## 1    1987    1988    1989    1990    1991    1992    1993    1994    1995
## 2 41,844  40,823  41,733  43,645  43,896  45,403  43,041  45,714  42,661
## 3 70,468  67,745  70,599  73,433  73,224  73,541  73,670  76,258  78,710
## 4 56,719  54,099  55,984  54,609  55,420  51,649  52,355  52,601  50,658
## 5 39,921  41,282  42,025  42,578  42,254  42,015  39,535  42,972  42,370
## 6 63,922  61,982  64,723  62,206  60,697  61,404  58,470  59,388  60,745
##      X.12    X.13    X.14    X.15    X.16    X.17    X.18    X.19    X.20
## 1    1996    1997    1998    1999    2000    2001    2002    2003    2004
## 2 48,447  49,983  55,997  54,806  51,798  49,989  52,633  50,988  48,812
## 3 84,383  75,109  78,271  77,703  77,275  81,556  73,868  70,945  73,378
## 4 50,581  51,237  57,269  55,931  58,172  60,715  55,616  56,340  58,430
## 5 43,364  40,942  42,716  44,875  43,424  47,400  45,332  43,798  46,620
## 6 62,052  62,120  63,204  65,961  68,456  67,195  66,397  67,473  65,594
##      X.21    X.22    X.23    X.24    X.25    X.26    X.27    X.28    X.29
## 1    2005    2006    2007    2008    2009    2010    2011    2012    2013
## 2 47,885  47,382  51,248  51,999  46,905  47,244  47,654  47,624  51,092
## 3 72,041  70,437  76,477  74,812  72,275  66,767  64,260  69,739  78,249
## 4 58,319  58,250  57,321  54,849  53,662  54,127  54,403  51,546  56,805
## 5 47,250  46,265  49,527  46,282  42,867  44,536  46,213  42,752  42,515
## 6 66,710  69,065  67,664  66,658  65,858  62,652  59,713  62,477  65,640
##      X.30    X.31    X.32    X.33    X.34               X.35
## 1    2014    2015    2016    2017    2018 % change 1984-2018
## 2 44,888  47,175  49,412  52,105  49,936             24.97%
## 3 71,804  79,611  79,237  79,888  68,734             -7.97%
## 4 52,294  55,378  59,750  61,155  62,283             25.93%
## 5 47,695  45,362  48,037  50,964  49,781             37.59%
## 6 64,221  67,448  69,729  71,745  70,489             20.76%
##                 X.36 X.37 X.38 X.39 X.40 X.41 X.42 X.43 X.44 X.45 X.46 X.47
## 1 Avg Annual % Change   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 2              0.81%   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 3             -0.04%   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 4              0.79%   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 5              1.08%   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 6              0.61%   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##   X.48 X.49 X.50 X.51 X.52 X.53 X.54 X.55 X.56 X.57 X.58 X.59 X.60 X.61 X.62
## 1   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 2   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 3   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 4   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 5   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## 6   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##   X.63 X.64 X.65 X.66 X.67
## 1   NA   NA   NA   NA   NA
## 2   NA   NA   NA   NA   NA
## 3   NA   NA   NA   NA   NA
## 4   NA   NA   NA   NA   NA
```

```
## 5   NA   NA   NA   NA   NA
## 6   NA   NA   NA   NA   NA
```

```
## cleaning the data set.
median_income1 <- median_income

colnames(median_income1) <- median_income[1, ]

median_income1 <- median_income1[-1, ]

colnames(median_income1)[1] <- "States"

main_median_income_ <- select(median_income1, "States", "2003", "2004", "2005", "2006", "2007", "2008",
                              "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016",)

# Cleaned up Median Income data set
head(main_median_income_)
```

```
##       States    2003    2004    2005    2006    2007    2008    2009    2010
## 2    Alabama 50,988  48,812  47,885  47,382  51,248  51,999  46,905  47,244
## 3     Alaska 70,945  73,378  72,041  70,437  76,477  74,812  72,275  66,767
## 4    Arizona 56,340  58,430  58,319  58,250  57,321  54,849  53,662  54,127
## 5   Arkansas 43,798  46,620  47,250  46,265  49,527  46,282  42,867  44,536
## 6 California 67,473  65,594  66,710  69,065  67,664  66,658  65,858  62,652
## 7   Colorado 68,349  67,811  65,026  69,537  74,228  71,251  65,618  69,520
##      2011    2012    2013    2014    2015    2016
## 2 47,654  47,624  51,092  44,888  47,175  49,412
## 3 64,260  69,739  78,249  71,804  79,611  79,237
## 4 54,403  51,546  56,805  52,294  55,378  59,750
## 5 46,213  42,752  42,515  47,695  45,362  48,037
## 6 59,713  62,477  65,640  64,221  67,448  69,729
## 7 65,601  62,735  73,325  64,702  70,585  73,841
```

```
# Second Data Set
## Load the Credit Card Debt data set
CCD <- read.csv("Credit Card Debt.csv")
## Raw Credit Card Data Set
head(CCD)
```

```
##   Citation..State.Level.Household.Debt.Statistics.2003.2022..Federal.Reserve.Bank.of.New.York..March
## 1
## 2                                                 Source: New York Fed Consumer Credit Panel / E
## 3
## 4
## 5
## 6
##           X       X.1       X.2       X.3       X.4       X.5       X.6
## 1
## 2
## 3  Q4_2003   Q4_2004   Q4_2005   Q4_2006   Q4_2007   Q4_2008   Q4_2009
## 4    4,260     4,430     4,440     4,510     4,850     4,810     4,540
## 5    2,600     2,600     2,510     2,560     2,790     2,820     2,510
## 6    2,430     2,460     2,520     2,540     2,720     2,730     2,390
```

```
##         X.7       X.8       X.9      X.10      X.11      X.12      X.13    X.14
## 1
## 2
## 3  Q4_2010   Q4_2011   Q4_2012   Q4_2013   Q4_2014   Q4_2015   Q4_2016  Q4_2017
## 4    4,180     4,060     3,980     3,820     3,830     3,920     4,110   4,270
## 5    2,320     2,240     2,180     2,050     2,070     2,110     2,190   2,370
## 6    2,180     2,120     2,060     1,970     1,990     2,050     2,150   2,300
##        X.15      X.16      X.17      X.18      X.19
## 1
## 2
## 3  Q4_2018   Q4_2019   Q4_2020   Q4_2021   Q4_2022
## 4    4,350     4,440     3,900     4,070     4,430
## 5    2,420     2,550     2,330     2,370     2,690
## 6    2,410     2,530     2,300     2,380     2,670
```

```
## Cleaning up the data set.
CCD1 <- CCD

CCD2 <- CCD1[-c(1:2), ]

colnames(CCD2) <- CCD1[c(3), ]

CCD3 <- CCD2[-1, ]

Main_CCD <- select(CCD3, c(1:15))

# Cleaned up Credit Card Data Set
head(Main_CCD)
```

```
##    state   Q4_2003   Q4_2004   Q4_2005   Q4_2006   Q4_2007   Q4_2008   Q4_2009
## 4     AK     4,260     4,430     4,440     4,510     4,850     4,810     4,540
## 5     AL     2,600     2,600     2,510     2,560     2,790     2,820     2,510
## 6     AR     2,430     2,460     2,520     2,540     2,720     2,730     2,390
## 7     AZ     3,110     3,170     2,970     3,100     3,590     3,990     3,740
## 8     CA     3,080     3,120     3,160     3,450     3,960     4,360     4,020
## 9     CO     3,480     3,650     3,730     3,880     4,240     4,450     4,090
##    Q4_2010   Q4_2011   Q4_2012   Q4_2013   Q4_2014   Q4_2015   Q4_2016
## 4    4,180     4,060     3,980     3,820     3,830     3,920     4,110
## 5    2,320     2,240     2,180     2,050     2,070     2,110     2,190
## 6    2,180     2,120     2,060     1,970     1,990     2,050     2,150
## 7    3,270     3,060     2,930     2,740     2,740     2,810     2,930
## 8    3,530     3,310     3,190     3,100     3,050     3,060     3,220
## 9    3,720     3,560     3,440     3,220     3,230     3,250     3,380
```

```
# Third Data Set
## Load the Total Household Debt data set
THHD <- read.csv("Total_Household_Debt.csv")
## Raw Total Debt Data Set
head(THHD)
```

```
##   Citation..State.Level.Household.Debt.Statistics.2003.2022..Federal.Reserve.Bank.of.New.York..March
## 1                                                                                                   
## 2                                                                  Source: New York Fed Consumer Credit Panel / E
```

```
## 3
## 4
## 5
## 6
##          X      X.1      X.2      X.3      X.4      X.5      X.6      X.7
## 1
## 2
## 3  Q4_2003  Q4_2004  Q4_2005  Q4_2006  Q4_2007  Q4_2008  Q4_2009  Q4_2010
## 4  36,420   40,530   43,290   46,450   51,490   54,110   53,140   52,730
## 5  24,390   26,390   26,600   29,530   33,240   35,540   34,180   33,590
## 6  20,260   21,510   23,250   24,550   27,480   28,940   27,990   27,470
##        X.8      X.9     X.10     X.11     X.12     X.13     X.14     X.15
## 1
## 2
## 3  Q4_2011  Q4_2012  Q4_2013  Q4_2014  Q4_2015  Q4_2016  Q4_2017  Q4_2018
## 4  48,440   54,740   54,000   54,330   56,050   57,040   57,850   58,550
## 5  34,390   34,370   33,380   33,760   34,210   34,780   36,060   36,780
## 6  28,910   29,480   29,030   29,430   30,020   30,820   32,220   32,790
##       X.16     X.17     X.18     X.19
## 1
## 2
## 3  Q4_2019  Q4_2020  Q4_2021  Q4_2022
## 4  59,630   60,400   63,920   67,670
## 5  38,000   39,140   41,220   44,440
## 6  33,950   35,180   37,000   39,150
```

```r
## Cleaning up the data set.
THHD1 <- THHD

THHD2 <- THHD1[-c(1:2), ]

colnames(THHD2) <- THHD1[c(3), ]

THHD3 <- THHD2[-1, ]

Main_THHD <- select(THHD3, c(1:15))

## Cleaned up Total Debt Data Set.
head(Main_THHD)
```

```
##   state  Q4_2003  Q4_2004  Q4_2005  Q4_2006  Q4_2007  Q4_2008  Q4_2009  Q4_2010
## 4    AK  36,420   40,530   43,290   46,450   51,490   54,110   53,140   52,730
## 5    AL  24,390   26,390   26,600   29,530   33,240   35,540   34,180   33,590
## 6    AR  20,260   21,510   23,250   24,550   27,480   28,940   27,990   27,470
## 7    AZ  37,720   41,950   46,190   55,550   63,750   65,870   60,200   54,320
## 8    CA  49,410   57,710   65,970   77,070   85,940   86,940   81,410   74,960
## 9    CO  50,010   55,230   57,890   62,090   66,140   68,830   65,490   61,640
##    Q4_2011  Q4_2012  Q4_2013  Q4_2014  Q4_2015  Q4_2016
## 4  48,440   54,740   54,000   54,330   56,050   57,040
## 5  34,390   34,370   33,380   33,760   34,210   34,780
## 6  28,910   29,480   29,030   29,430   30,020   30,820
## 7  52,210   48,890   46,950   47,230   47,630   48,590
## 8  72,320   68,970   65,440   66,140   65,740   66,970
## 9  62,390   60,830   59,870   60,900   62,200   64,090
```

```r
# Creating a new Date frame.
## US Median Income
US_MI <- main_median_income_[52, ]
US_MI1 <- US_MI[,-1]

## US Credit Card Debt
US_CCD <- Main_CCD[53, ]
US_CCD1 <- US_CCD[,-1]
setnames(US_CCD1, new = c("2003", "2004", "2005", "2006",
                          "2007", "2008", "2009", "2010", "2011", "2012", "2013", "2014",
                          "2015", "2016"))

## US Total Debt
US_TD <- Main_THHD[53,]
US_TD1 <- US_TD[,-1]
setnames(US_TD1, new = c("2003", "2004", "2005", "2006",
                          "2007", "2008", "2009", "2010", "2011", "2012", "2013", "2014",
                          "2015", "2016"))

## Combining the US Median Income, Credit Card Debt and Total Debt into one df
project_data <- bind_rows(US_MI1, US_TD1, US_CCD1)
print(project_data)
```

```
##       2003     2004     2005     2006     2007     2008     2009     2010
## 1  59,286   59,080   59,712   60,178   60,985   58,811   58,400   56,873
## 2  33,430   37,290   40,650   45,410   50,170   52,010   49,820   47,410
## 3   2,960    3,040    3,060    3,170    3,490    3,670    3,370    3,050
##       2011     2012     2013     2014     2015     2016
## 1  56,006   55,900   57,856   56,969   59,901   61,779
## 2  47,790   47,020   45,310   45,710   46,000   46,950
## 3   2,950    2,850    2,710    2,730    2,800    2,930
```

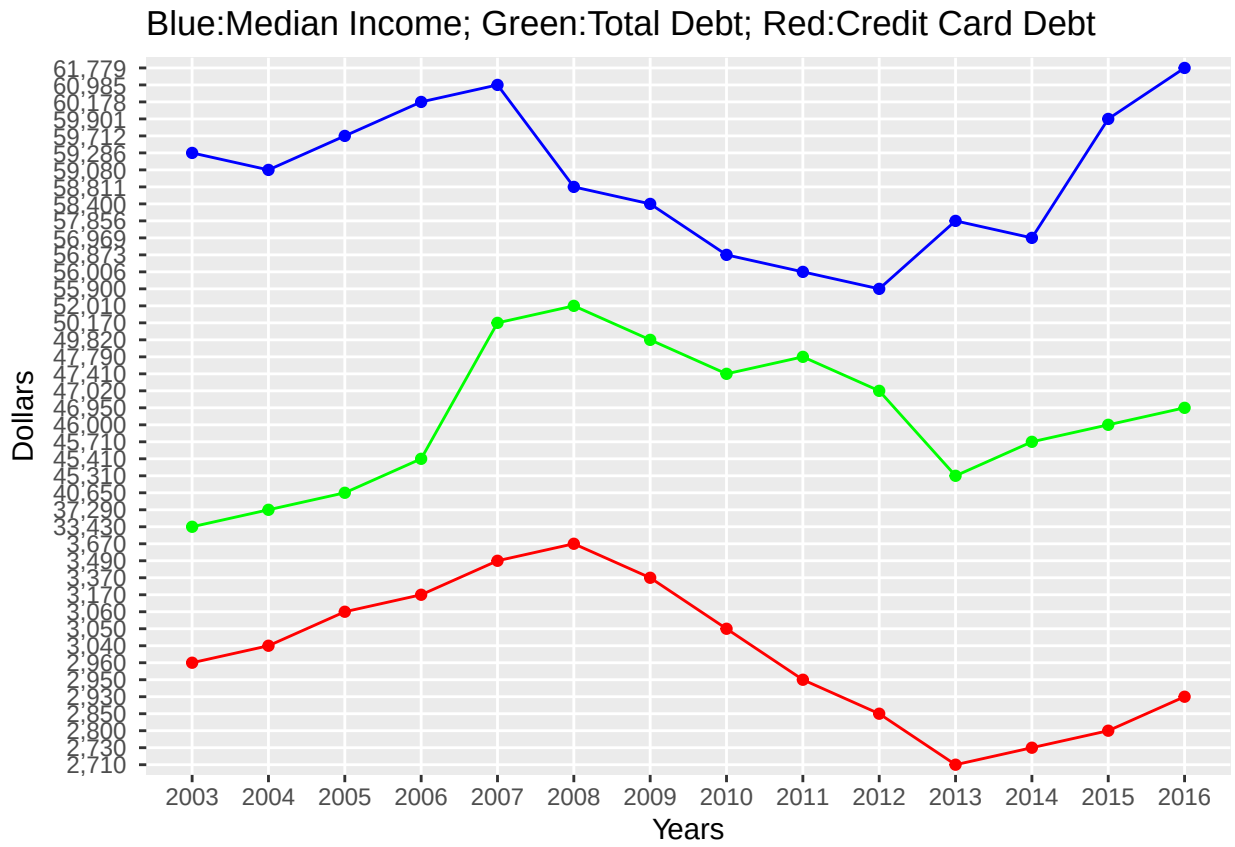
```r
new_project_data <- t(project_data)
print(new_project_data)
```

```
##      [,1]       [,2]       [,3]
## 2003 "59,286 " " 33,430 " " 2,960 "
## 2004 "59,080 " " 37,290 " " 3,040 "
## 2005 "59,712 " " 40,650 " " 3,060 "
## 2006 "60,178 " " 45,410 " " 3,170 "
## 2007 "60,985 " " 50,170 " " 3,490 "
## 2008 "58,811 " " 52,010 " " 3,670 "
## 2009 "58,400 " " 49,820 " " 3,370 "
## 2010 "56,873 " " 47,410 " " 3,050 "
## 2011 "56,006 " " 47,790 " " 2,950 "
## 2012 "55,900 " " 47,020 " " 2,850 "
## 2013 "57,856 " " 45,310 " " 2,710 "
## 2014 "56,969 " " 45,710 " " 2,730 "
## 2015 "59,901 " " 46,000 " " 2,800 "
## 2016 "61,779 " " 46,950 " " 2,930 "
```

```r
project_df <- data.frame(new_project_data)
setnames(project_df, new = c("Median_Income", "Total_Debt", "Credit_Card_Debt"))
print(project_df)
```

```
##      Median_Income Total_Debt Credit_Card_Debt
## 2003        59,286     33,430            2,960
## 2004        59,080     37,290            3,040
## 2005        59,712     40,650            3,060
## 2006        60,178     45,410            3,170
## 2007        60,985     50,170            3,490
## 2008        58,811     52,010            3,670
## 2009        58,400     49,820            3,370
## 2010        56,873     47,410            3,050
## 2011        56,006     47,790            2,950
## 2012        55,900     47,020            2,850
## 2013        57,856     45,310            2,710
## 2014        56,969     45,710            2,730
## 2015        59,901     46,000            2,800
## 2016        61,779     46,950            2,930
```

```r
## Creating a visual for the data
ggplot() +
  geom_line(data = project_df, mapping = aes( x=(rownames(project_df)),
  y=Median_Income), group = 14, color = "blue") +
  geom_point(data = project_df, mapping = aes(x=(rownames(project_df)),
  y=Median_Income), group = 14, color = "blue") +
  geom_line(data = project_df, mapping = aes(x=(rownames(project_df)),
  y=Total_Debt), group = 14, color = "green") +
  geom_point(data = project_df, mapping = aes(x=(rownames(project_df)),
  y=Total_Debt), group = 14, color = "green") +
  geom_line(data = project_df, mapping = aes(x=(rownames(project_df)),
  y=Credit_Card_Debt), group = 14, color = "red") +
  geom_point(data = project_df, mapping = aes(x=(rownames(project_df)),
  y=Credit_Card_Debt), group = 14, color = "red") +
  xlab("Years") + ylab("Dollars") + ggtitle("Blue:Median Income; Green:Total Debt; Red:Credit Card Debt
```




```
## Calculating multiple regression
#### I couldn't get the code below to work
#### project_lm <- lm(Median_Income ~ Total_Debt + Credit_Card_Debt, data = project_df)
```