problemSet_2

Tianjian Xie

2022-09-22

Fuel Economy Part1

1. Do an analysis of Fuel economy over the 40 year span 1984 through 2023, inclusive. You may want to do an analysis by type of fuel which will ignore hybrids and electric vehicles for most the the years under analysis.

```
vehicles <- read.csv("C:/Users/JasonXie/Desktop/TianjianXie/BU/MSSP/MA615_Data Science in R/HW/Set 2/vehicles.csv")
del <- c()
for(i in 1:length(vehicles$fuelType)){
  if(grepl('Electricity',vehicles$fuelType[i])){
    del <- c(del,i)
  }
}
vehicles <- vehicles[-del,]
unique(vehicles$fuelType)
```

```
## [1] "Regular"                "Premium"
## [3] "Diesel"                 "CNG"
## [5] "Gasoline or natural gas" "Gasoline or E85"
## [7] "Gasoline or propane"    "Premium or E85"
## [9] "Midgrade"
```

```
table(vehicles$fuelType)
```

```
##
##                     CNG                  Diesel         Gasoline or E85
##                      60                    1253                    1377
## Gasoline or natural gas     Gasoline or propane                Midgrade
##                      20                       8                     148
##                 Premium          Premium or E85                 Regular
##                   13625                     127                   28179
```

Fuel Economy Part2

2. Now, examine vehicle makers. Which ones have made the most progress? Make at least two plots that address the questions above. As you do your work, you may make many plots. If you include plots in addition to the two that described above, make sure that they address different issues and are not simply intermediate steps you took as you made the to plots you're submitting for questions 1 and 2.

```
vehicles2 <- read.csv("C:/Users/JasonXie/Desktop/TianjianXie/BU/MSSP/MA615_Data Science in R/HW/Set2/vehicles.csv")
#Find which car maker has large sample
table(vehicles2$make)
```

```
##
##                          Acura                      Alfa Romeo
##                            388                              87
##                     AM General     American Motors Corporation
##                              6                              27
##                ASC Incorporated                   Aston Martin
##                              1                             176
##                           Audi                 Aurora Cars Ltd
##                           1199                               1
##              Autokraft Limited        Avanti Motor Corporation
##                              4                               2
##                Azure Dynamics                         Bentley
##                              2                             163
##                        Bertone    Bill Dovell Motor Car Company
##                              7                               4
##          Bitter Gmbh and Co. Kg                             BMW
##                              5                            2296
##                    BMW Alpina                         Bugatti
##                              3                              19
##                          Buick                             BYD
##                            705                               7
##                       Cadillac                 CCC Engineering
##                            709                               2
##                      Chevrolet                        Chrysler
##                           4330                             747
##                CODA Automotive        Consulier Industries Inc
##                              2                               3
##                 CX Automotive      Dabryan Coach Builders Inc
##                             17                               9
##                          Dacia                           Daewoo
##                              3                              67
##                       Daihatsu                           Dodge
##                             17                            2659
##             E. P. Dutton, Inc.                           Eagle
##                              1                             161
## Environmental Rsch and Devp Corp              Evans Automobiles
##                              1                               3
##                Excalibur Autos                  Federal Coach
##                              1                              14
##                        Ferrari                           Fiat
##                            260                              77
##                         Fisker                           Ford
##                              1                            3644
##                General Motors                        Genesis
##                              1                              96
##                           Geo                            GMC
##                            147                            2724
##                       Goldacre        Grumman Allied Industries
##                              1                               1
##                 Grumman Olson                          Honda
##                              4                            1111
##                        Hummer                        Hyundai
```

```
##                              19                             918
##      Import Foreign Auto Sales Inc             Import Trade Services
##                               1                              13
##                        Infiniti                 Isis Imports Ltd
##                             463                               1
##                           Isuzu                     J.K. Motors
##                             434                              36
##                          Jaguar                JBA Motorcars, Inc.
##                             521                               1
##                            Jeep                           Kandi
##                            1068                               1
##                           Karma      Kenyon Corporation Of America
##                               5                               4
##                             Kia                       Koenigsegg
##                             734                               3
##            Laforza Automobile Inc            Lambda Control Systems
##                               2                               1
##                      Lamborghini                      Land Rover
##                             155                             307
##                           Lexus                          Lincoln
##                             610                             397
##               London Coach Co Inc                      London Taxi
##                               1                               1
##                           Lotus                           Lucid
##                              65                               6
##                        Mahindra                        Maserati
##                               1                             187
##                         Maybach                           Mazda
##                              31                            1079
##                    Mcevoy Motors                McLaren Automotive
##                               6                              48
##                    Mercedes-Benz                         Mercury
##                            1766                             609
##                          Merkur                            MINI
##                              14                             502
##                      Mitsubishi             Mobility Ventures LLC
##                            1115                               4
##                          Morgan                          Nissan
##                               3                            1594
##                       Oldsmobile                         Pagani
##                             462                               4
##                           Panos             Panoz Auto-Development
##                               1                               1
##        Panther Car Company Limited                    PAS Inc - GMC
##                               4                               2
##                         PAS, Inc                         Peugeot
##                               2                              98
##                      Pininfarina                        Plymouth
##                               6                             526
##                         Polestar                         Pontiac
##                               9                             893
##                          Porsche               Quantum Technologies
```

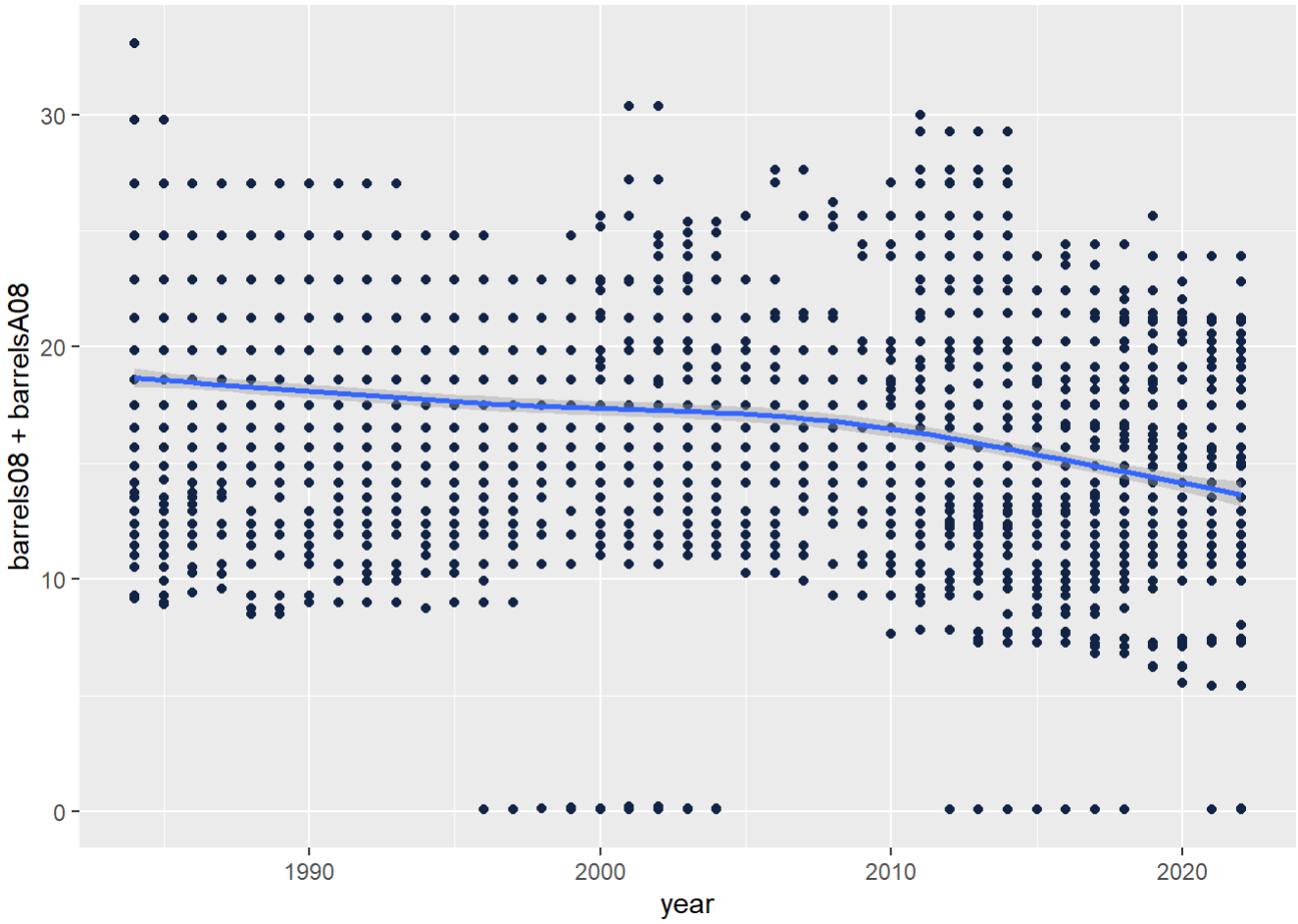
```
##                             1318                                2
##                            Qvale                              Ram
##                                1                              127
##                    Red Shift Ltd.                          Renault
##                                2                               56
##                           Rivian                      Rolls-Royce
##                                2                              221
##                  Roush Performance                  RUF Automobile
##                               69                                3
##              Ruf Automobile Gmbh S and S Coach Company  E.p. Dutton
##                                3                                1
##                             Saab                           Saleen
##                              432                                5
##                 Saleen Performance                          Saturn
##                                5                              278
##                            Scion                          Shelby
##                               84                                1
##                            smart                          Spyker
##                               38                               13
##                              SRT                        Sterling
##                                2                               12
##                              STI                          Subaru
##                                1                              972
##  Superior Coaches Div E.p. Dutton                          Suzuki
##                                1                              515
##                      Tecstar, LP                           Tesla
##                                6                              124
##              Texas Coach Company                          Toyota
##                                4                             2296
##              TVR Engineering Ltd                          Vector
##                                4                                4
##              Vixen Motor Company       Volga Associated Automobile
##                                1                                1
##                       Volkswagen                          Volvo
##                             1280                              893
##                              VPG            Wallace Environmental
##                                5                               32
##                             Yugo
##                                8
```

```
#Select Ford(3644 Samples),Chevrolet(4330 Samples), and Dodge(2659 Samples) to make comparison.
Fordtable <- vehicles2[vehicles2$make == "Ford", ]
Chevrolettable <- vehicles2[vehicles2$make == "Chevrolet", ]
Dodgetable <- vehicles2[vehicles2$make == "Dodge", ]
#To explain what does "progress" mean, I will choose barrels08(annual petroleum consumption in bar
rels for fuelType1 (1)), barrelsA08(annual petroleum consumption in barrels for fuelType2 (1)) as
 year going
plotFord <- ggplot(Fordtable) +
  aes(x = year, y = barrels08 + barrelsA08) +
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  geom_smooth(span = 0.75)
  labs(
    x = "Year",
    y = "Barrels",
    title = "Ford Cars' Annual Petroleum Consumption in Barrels",
    subtitle = "1984-2023"
  ) +
  theme_minimal()
```

```
## NULL
```

```
  plotFord
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

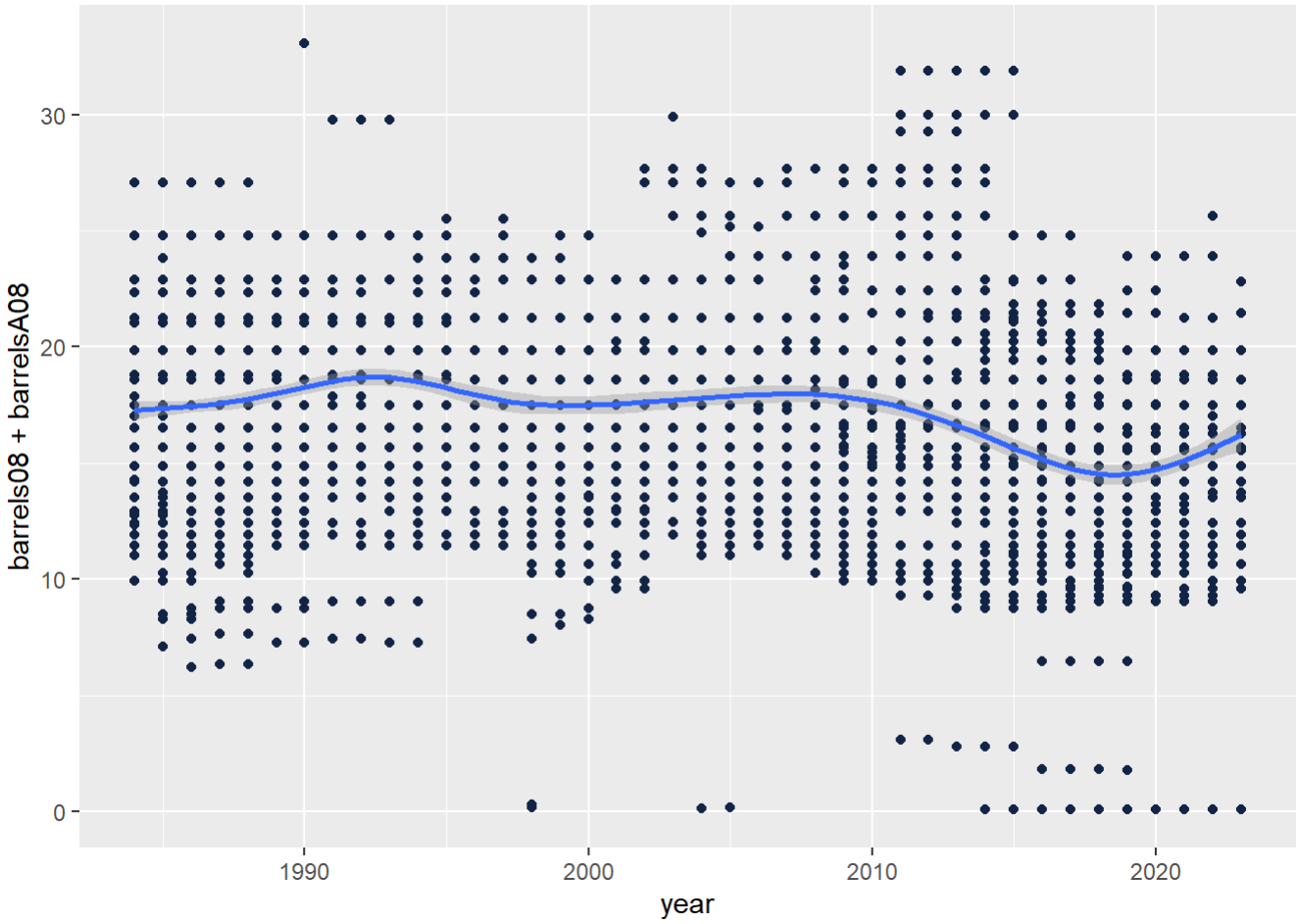



```
plotChevrolet <- ggplot(Chevrolettable) +
  aes(x = year, y = barrels08 + barrelsA08) +
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  geom_smooth(span = 0.75)
  labs(
    x = "Year",
    y = "Barrels",
    title = "Chevrolet Cars' Annual Petroleum Consumption in Barrels",
    subtitle = "1984-2023"
  ) +
  theme_minimal()
```

```
## NULL
```

```
  plotChevrolet
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

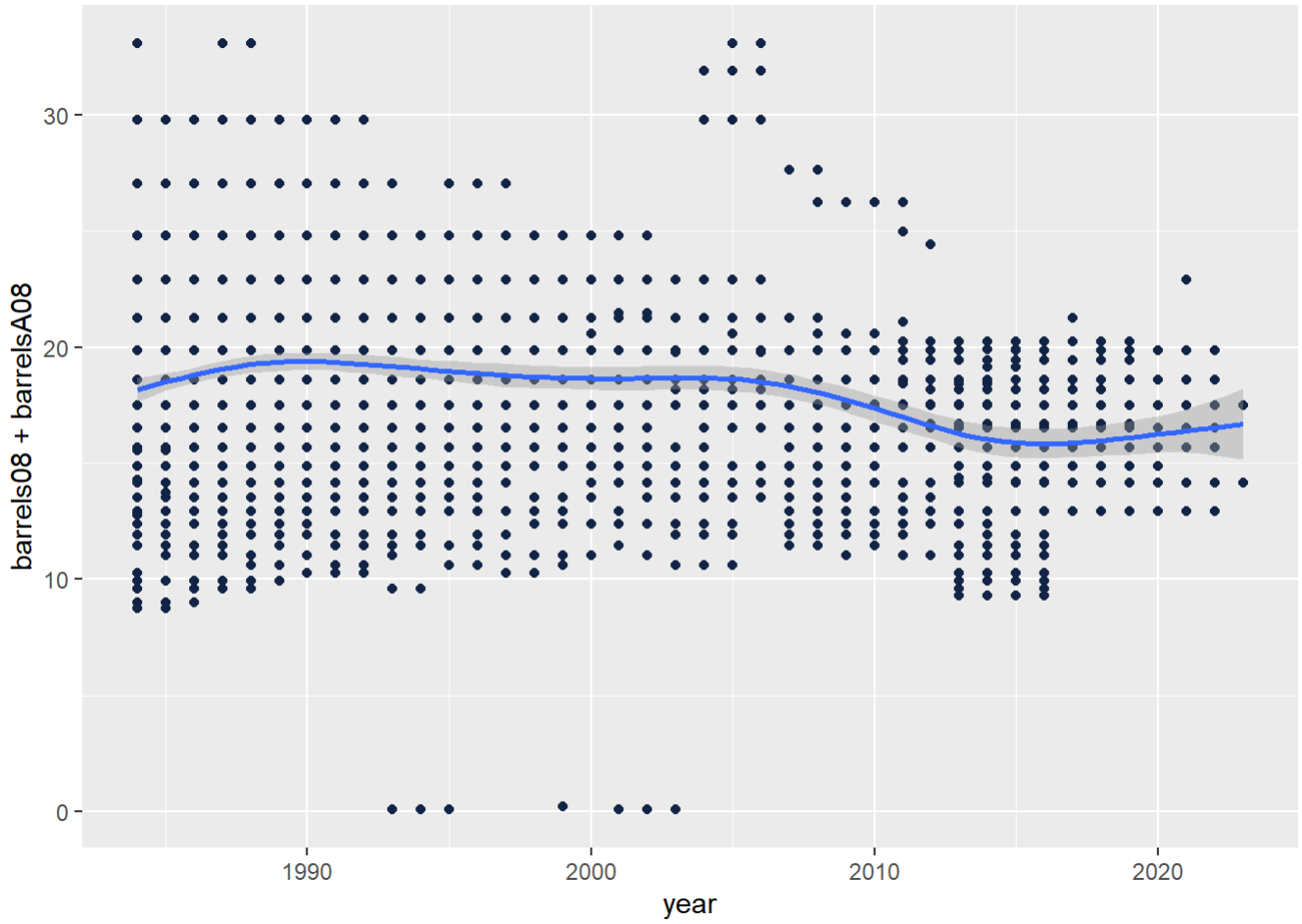



```
plotDodge <- ggplot(Dodgetable) +
  aes(x = year, y = barrels08 + barrelsA08) +
  geom_point(shape = "circle", size = 1.5, colour = "#112446") +
  geom_smooth(span = 0.75)
  labs(
    x = "Year",
    y = "Barrels",
    title = "Dodge Cars' Annual Petroleum Consumption in Barrels",
    subtitle = "1984-2023"
  ) +
  theme_minimal()
```

```
## NULL
```

```
  plotDodge
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
#From the graph we can see cars made by Ford have the largest progress on decrease the consumption of petroleum
```

NASDAQ Composite

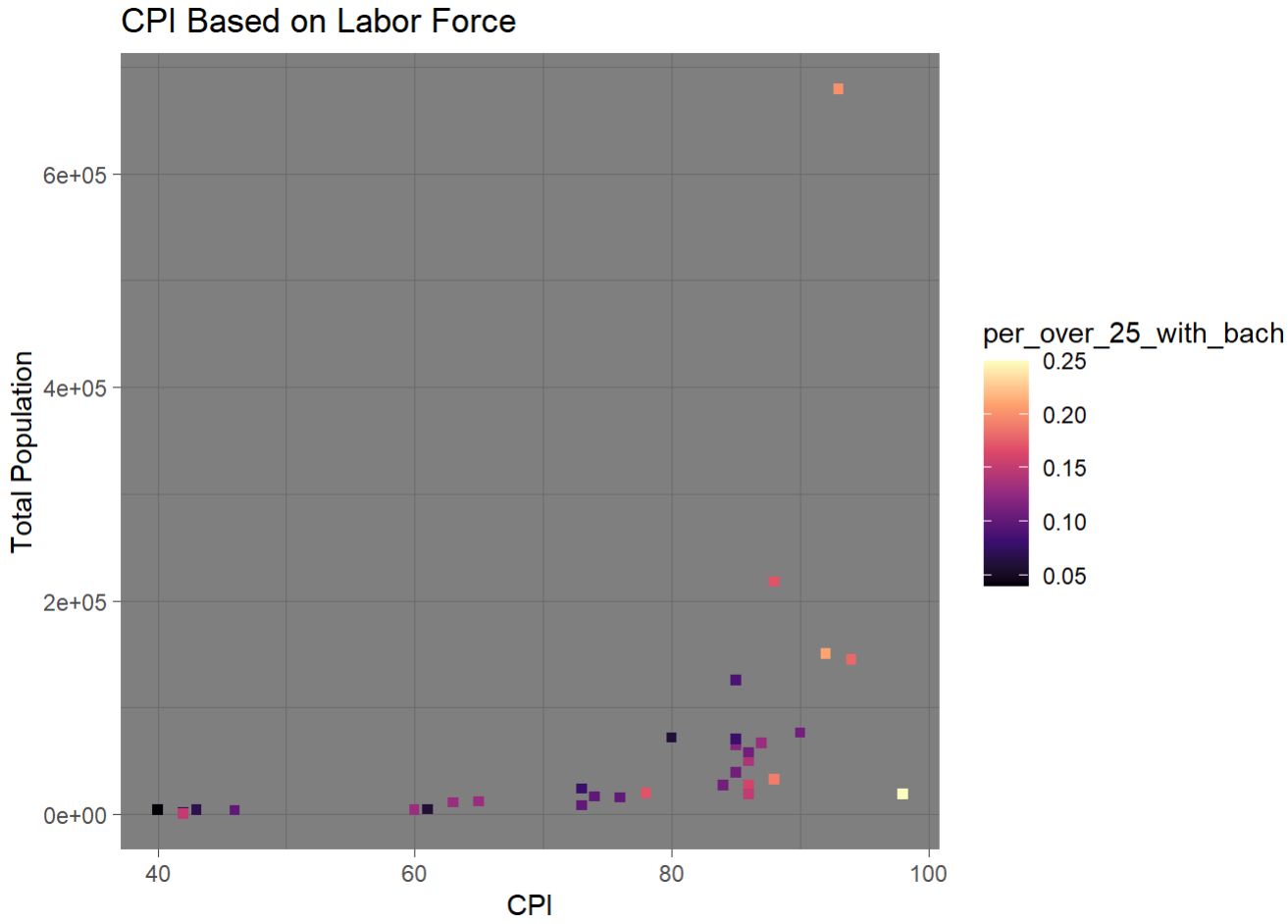
The Nasdaq Composite (ticker symbol ^IXIC) is a stock market index that includes almost all stocks listed on the Nasdaq stock exchange. Along with the Dow Jones Industrial Average and S&P 500, it is one of the three most-followed stock market indices in the United States. Your mission is using ggplot to create a Candlestick chart with the Nasdaq Composite data from September 20, 2021 to September 20, 2022, using file IXIC21-22.csv. Be aware of the following when you create your chart: • Make sure the X-axis is Date and the Y-axis is Adjust Close. • Include an appropriate title with your graph. • Make sure your data is clean before you create the plot. • The format for X label will be month-year or year-month eg, Jul-2022, 07-2022, 2022-07 • You might need to use tidyquant package to create this plot. • In a sentence or two, what does this plot show?

```
NASDAQdata <- read.csv("C:/Users/JasonXie/Desktop/TianjianXie/BU/MSSP/MA615_Data Science in R/HW/Set2/IXIC21-22.csv")
NASDAQdata$Date <- as.Date(NASDAQdata$Date)
ggplot(data = NASDAQdata, aes(x = Date, y = Adj.Close)) +
    geom_candlestick(aes(open = Open, high = High, low = Low, close = Close),colour_up = "darkgreen",colour_down = "red",fill_up = "darkgreen",fill_down = "red") +
    labs(title = "NASDAQ", y = "Adjust Close", x = "Date") +
    theme_tq()
```


```
#This shows the daily changing of Adjust Price, which means the amends closing price of stocks ref
lected the stocks' value after accounting for any corporate actions, of the NASDAQ, from 2021-09-2
1 to 2022-09-20
```

##Rural Capacity Index 1. Create a plot that emphasizes rural capacity indexes. Choose your other variables to reflect their contribution to the rural capacity index.

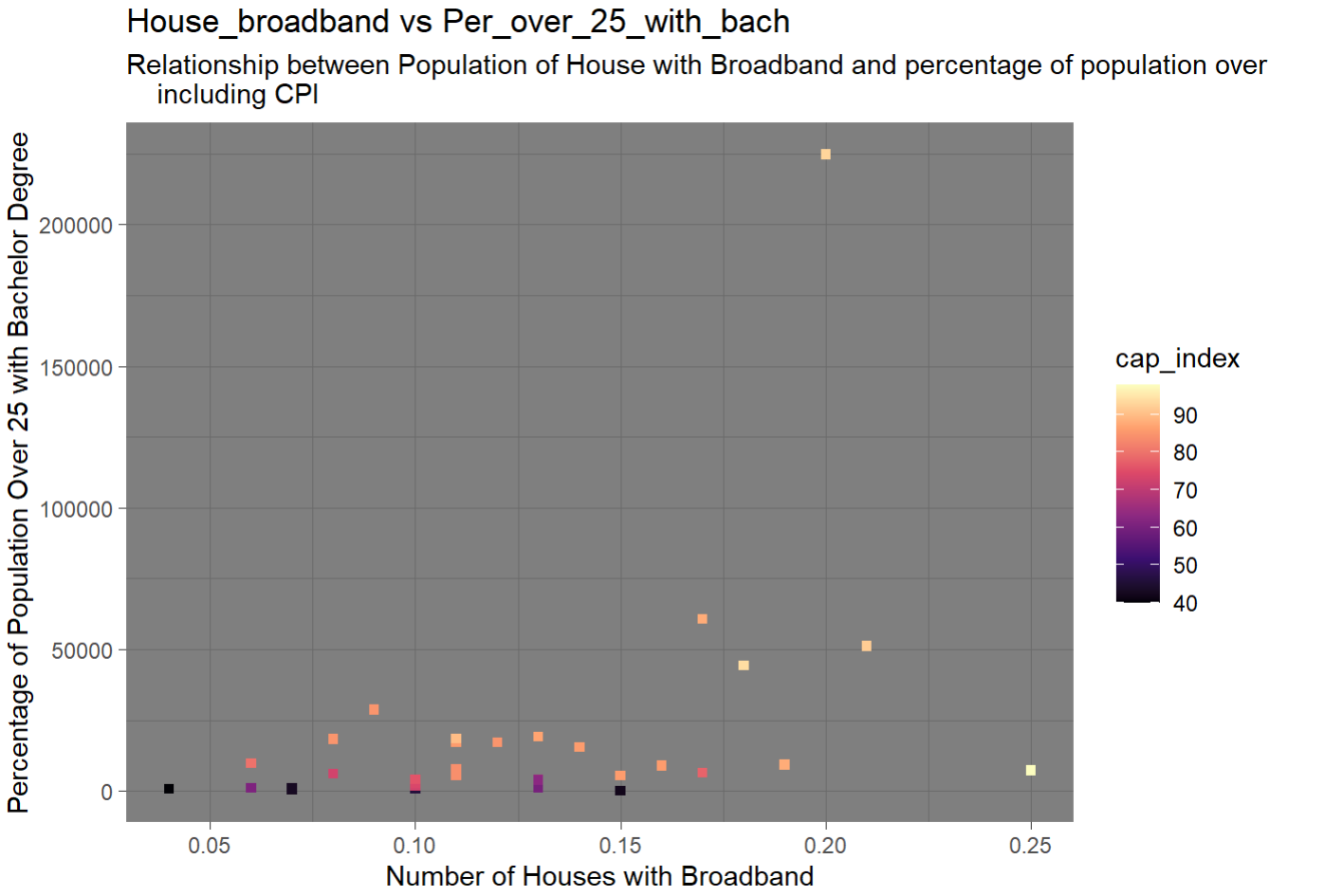
```
RCIdata <- read.csv("C:/Users/JasonXie/Desktop/TianjianXie/BU/MSSP/MA615_Data Science in R/HW/Set 2/ruralCapacityData.csv")
#1 I choose pop_total as y-axis and use per_over_25_with_bach to color the plot. This graph shows  that how the total population and the percentage of the population that can fit working affects t he CPI.
ggplot(RCIdata) +
  aes(
    x = cap_index,
    y = pop_total,
    colour = per_over_25_with_bach
  ) +
  geom_point(shape = "square", size = 1.5) +
  scale_color_viridis_c(option = "magma", direction = 1) +
  labs(
    x = "CPI",
    y = "Total Population",
    title = "CPI Based on Labor Force"
  ) +
  theme_dark()
```



#The graph shows that as the population of labor force(In this case, explained as the product of t he population and percentage of population over 25 with bachelor degree.) smaller, the CPI will be lower. If the population, or the percentage, or both variables are high, the CPI might also high.

##Rural Capacity Index 2. Create a plot that demonstrates the relationship between the number of houses with broadband and the percent of adults, 25 and older, with bachelor degrees. Include information about the rural capacity indexes.

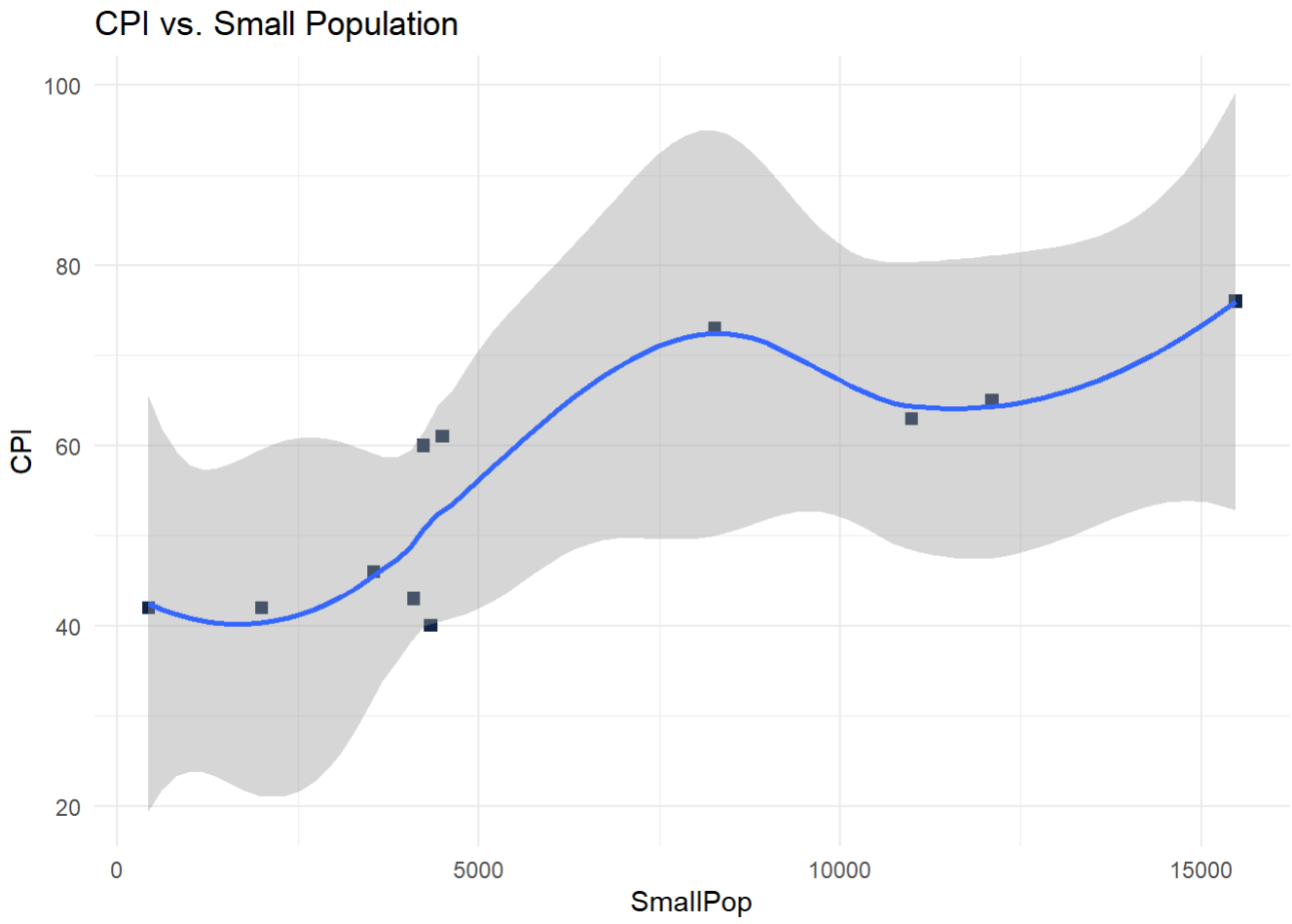
```
ggplot(RCIdata) +
  aes(
    x = per_over_25_with_bach,
    y = house_broadband,
    colour = cap_index
  ) +
  geom_point(shape = "square", size = 1.5) +
  scale_color_viridis_c(option = "magma", direction = 1) +
  labs(
    x = "Number of Houses with Broadband",
    y = "Percentage of Population Over 25 with Bachelor Degree",
    title = "House_broadband vs Per_over_25_with_bach",
    subtitle = "Relationship between Population of House with Broadband and percentage of populati
on over 25 with Bachelor degree,
    including CPI "
  ) +
  theme_dark()
```



##Rural Capacity Index 3. Explore different sizes of communities and their capacity indexes. Create three plots that describe communities with total population < 16000, 16000 < total population < 55000, and total population > 55000. What facets of each population subsection stand out to you, demonstrate them in your plots.

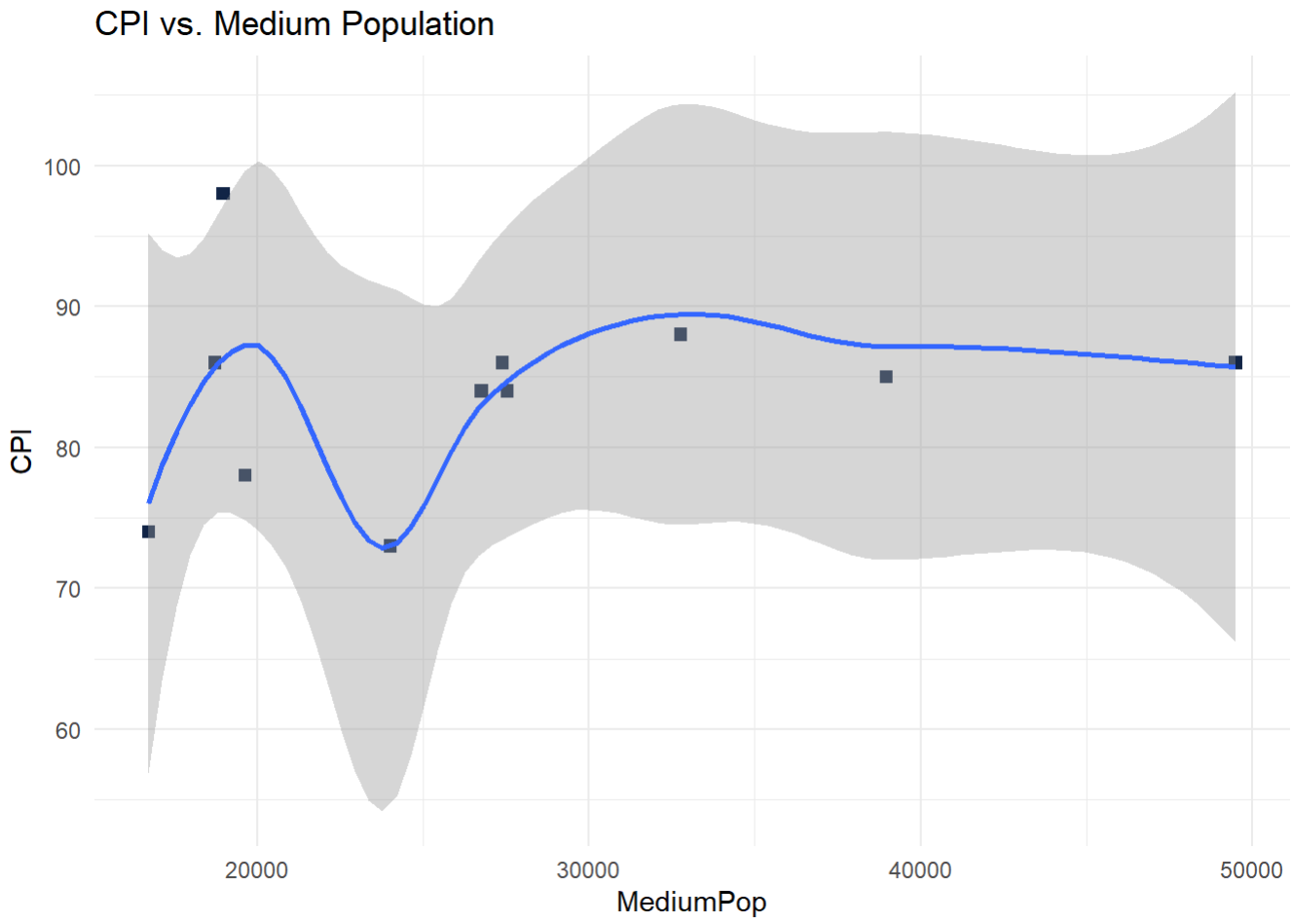
```
#Total Population < 16000
RCIdata %>%
 filter(pop_total < 16000L) %>%
 ggplot() +
  aes(x = pop_total, y = cap_index) +
  geom_point(shape = "square", size = 2L, colour = "#112446") +
  geom_smooth(span = 0.75) +
  labs(
    x = "SmallPop",
    y = "CPI",
    title = "CPI vs. Small Population"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
#Total Population between 16000 and 55000
RCIdata %>%
 filter(pop_total > 16000L & pop_total < 55000L) %>%
 ggplot() +
  aes(x = pop_total, y = cap_index) +
  geom_point(shape = "square", size = 2L, colour = "#112446") +
  geom_smooth(span = 0.75) +
  labs(
    x = "MediumPop",
    y = "CPI",
    title = "CPI vs. Medium Population"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
#Total Population > 55000
RCIdata %>%
 filter(pop_total > 55000L) %>%
 ggplot() +
  aes(x = pop_total, y = cap_index) +
  geom_point(shape = "square", size = 2L, colour = "#112446") +
  geom_smooth(span = 0.75) +
  labs(
    x = "LargePop",
    y = "CPI",
    title = "CPI vs. Large Population"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```