Markdown

2022-09-27

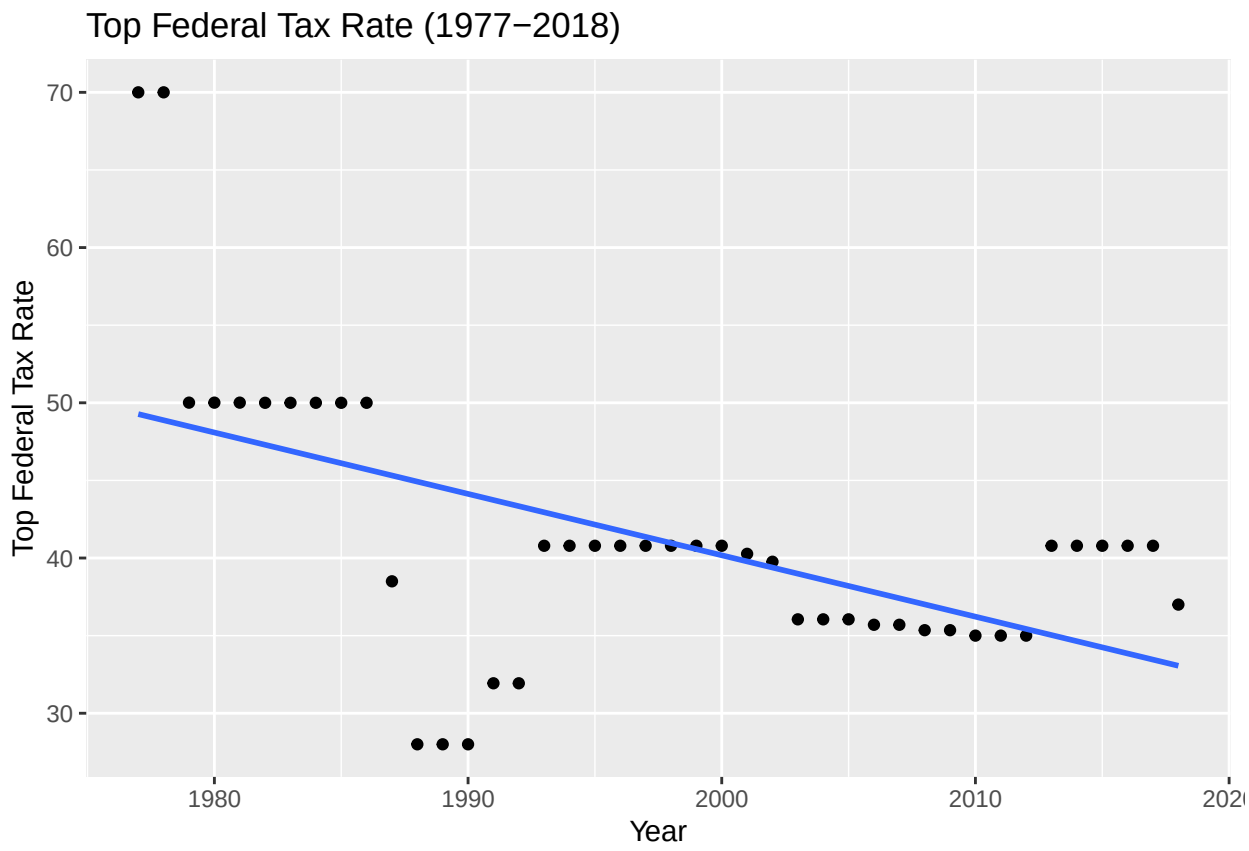
Question 1

The goal of this exercise is to estimate the mobility of high income US taxpayers across US states due to variation in state income top tax rates across states and over time. High income US taxpayers are defined as tax filers reporting Adjusted Gross Income (AGI) above $1m amongst all the other millionaires in a given year (add up all the millionaires in all the states + DC in a given year and compute the share in each state in a given year).

1. The file toprates 77 18.csv documents the top state and federal income tax rate each year from 1977-2018. Plot the the mean and median top state and federal tax rate across states is over the time-period 1977-2018. Are top rates trending upward or downward?

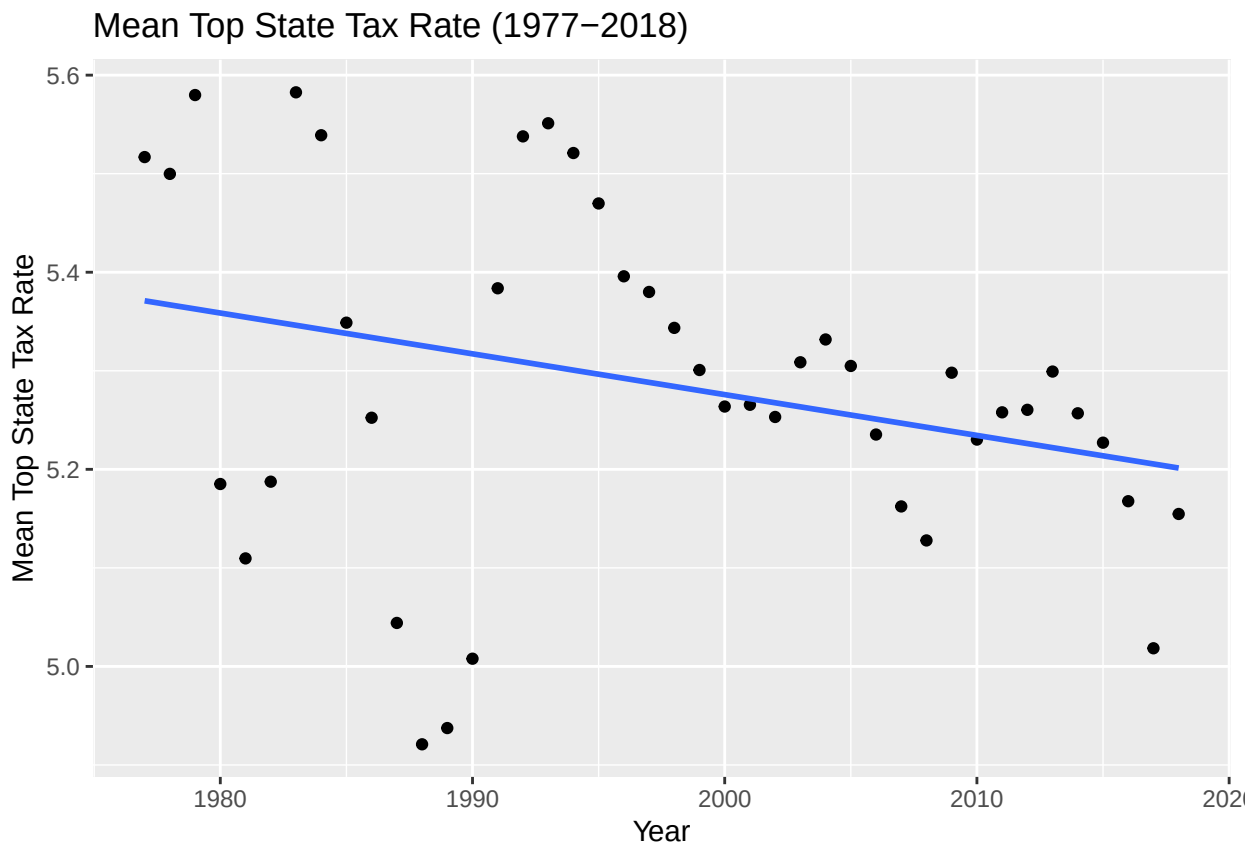
```
means_and_median_federal <- top_rates %>%
  filter(state_full == "FEDERAL") %>%
  group_by(year) %>%
  summarize(mean_federal=mean(topfederalrate))

means_and_median_state <- top_rates %>%
  filter(state_full != "FEDERAL") %>%
  group_by(year) %>%
  summarize(mean_state=mean(topstaterate),
            median_state=median(topstaterate))

ggplot(data = means_and_median_federal) + geom_point(aes(x = year, y = mean_federal)) +
 geom_smooth(aes(x = year, y = mean_federal), method = "lm", se = F) +
 labs(title = "Top Federal Tax Rate (1977-2018)",
      x = "Year",
      y = "Top Federal Tax Rate")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



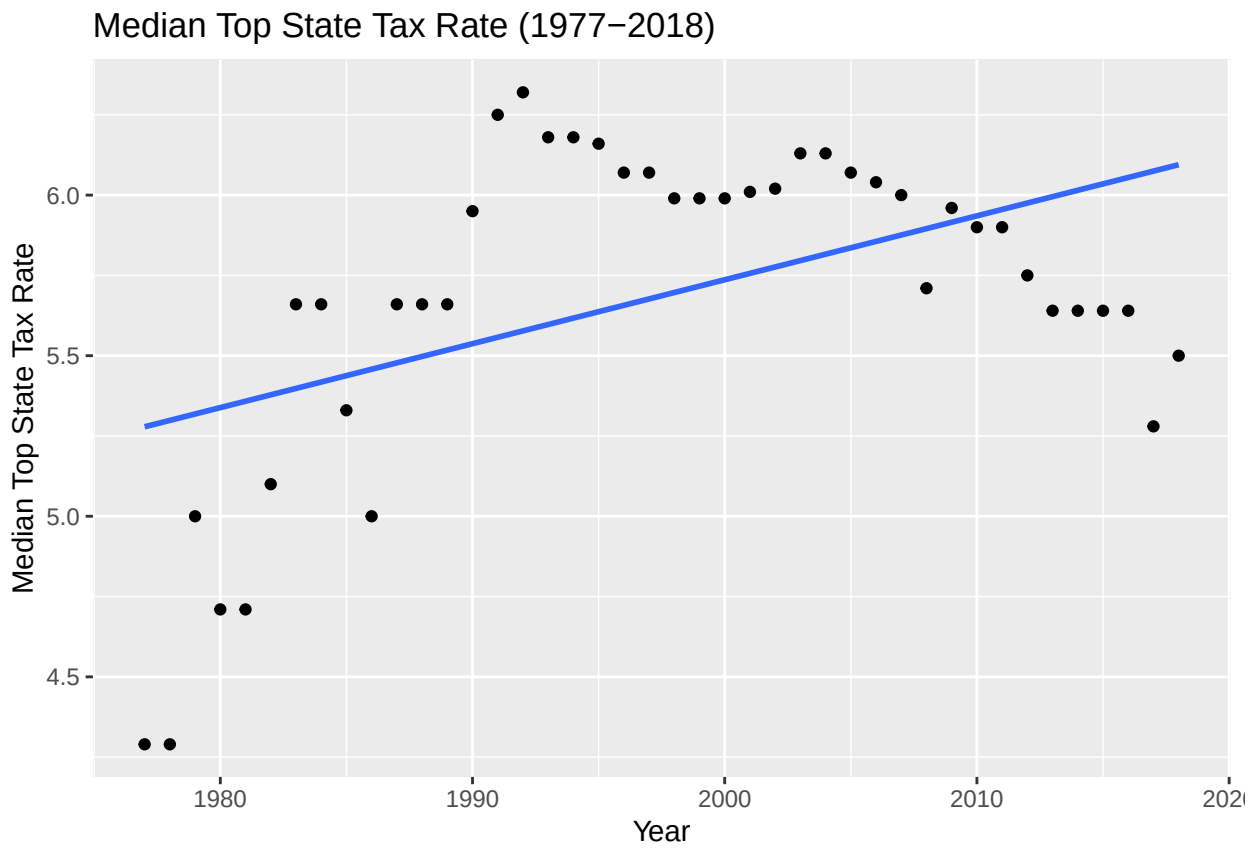

```
ggplot(data = means_and_median_state) +
  geom_point(aes(x = year, y = mean_state)) +
  geom_smooth(aes(x = year, y = mean_state), method = "lm", se = F) +
  labs(title = "Mean Top State Tax Rate (1977-2018)",
       x = "Year",
       y = "Mean Top State Tax Rate")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(data = means_and_median_state) +
  geom_point(aes(x = year, y = median_state)) +
  geom_smooth(aes(x = year, y = median_state), method = "lm", se = F) +
  labs(title = "Median Top State Tax Rate (1977-2018)",
       x = "Year",
       y = "Median Top State Tax Rate")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Top federal rates have trended down, as have the mean top state rates, but the median top state rates have increased.

```
group_t <- top_rates %>%
  filter(year == 2016) %>%
  filter(state_full != "FEDERAL") %>%
  arrange(topstaterate) %>%
  slice(unique(c(n() - 0:9)) ) %>%
  select(state_full, topstaterate)

group_c <- top_rates %>%
  filter(year == 2016) %>%
  filter(state_full != "FEDERAL") %>%
  arrange(topstaterate) %>%
  slice(unique(c(1:10)) ) %>%
  select(state_full, topstaterate)

group_t
```

```
## # A tibble: 10 x 2
##    state_full           topstaterate
##    <chr>                       <dbl>
##  1 CALIFORNIA                  14.1
##  2 MINNESOTA                   10.2
##  3 OREGON                       9.99
##  4 NEW JERSEY                   8.97
##  5 VERMONT                      8.95
```

```
##  6 DISTRICT OF COLUMBIA         8.84
##  7 HAWAII                       8.5
##  8 NEBRASKA                     7.71
##  9 WISCONSIN                    7.65
## 10 IDAHO                        7.49
```

```
group_c
```

```
## # A tibble: 10 x 2
##    state_full    topstaterate
##    <chr>                <dbl>
##  1 ALASKA                 0
##  2 FLORIDA                0
##  3 NEVADA                 0
##  4 NEW HAMPSHIRE          0
##  5 SOUTH DAKOTA           0
##  6 TENNESSEE              0
##  7 TEXAS                  0
##  8 WASHINGTON             0
##  9 WYOMING                0
## 10 NORTH DAKOTA           2.9
```

```
tax_rate_merged %>%
  filter(state_full %in% group_t$state_full & year == 2016) %>%
  summarize(mean(millionaire_share))
```

```
## # A tibble: 1 x 1
##   `mean(millionaire_share)`
##                       <dbl>
## 1                    0.0270
```

```
tax_rate_merged %>%
  filter(state_full %in% group_c$state_full & year == 2016) %>%
  summarize(mean(millionaire_share))
```

```
## # A tibble: 1 x 1
##   `mean(millionaire_share)`
##                       <dbl>
## 1                    0.0204
```

```
changes_temp <- tax_rate_merged %>%
  filter(year == 2001 | year == 2016) %>%
  select(state_full, topstaterate, year) %>%
  group_by(state_full) %>%
  pivot_wider(id_cols = state_full, names_from = year,
              values_from = c(topstaterate)) %>%
  rename("rate_2001" = 2) %>%
  rename("rate_2016" = 3) %>%
  mutate(diff = rate_2016 - rate_2001) %>%
  mutate(diff_pct = (rate_2016 - rate_2001)/rate_2001 * 100) %>%
  replace(is.na(.), 0)

group_tt <- changes_temp %>%
  arrange(desc(diff))  %>%
  head(10)

group_cc <- changes_temp %>%
```

```
  arrange(diff)  %>%
  head(10)

group_tt
```

```
## # A tibble: 10 x 5
## # Groups:   state_full [10]
##    state_full  rate_2001 rate_2016  diff diff_pct
##    <chr>           <dbl>     <dbl> <dbl>    <dbl>
##  1 CALIFORNIA       9.86     14.1   4.24     43.0
##  2 NEW JERSEY       6.37      8.97  2.6      40.8
##  3 CONNECTICUT      4.5       6.99  2.49     55.3
##  4 MINNESOTA        8.09     10.2   2.06     25.5
##  5 MARYLAND         4.87      5.83  0.96     19.7
##  6 OREGON           9.08      9.99  0.91     10.0
##  7 WISCONSIN        6.75      7.65  0.9      13.3
##  8 NEBRASKA         6.85      7.71  0.86     12.6
##  9 ILLINOIS         3         3.75  0.75     25
## 10 DELAWARE         6.13      6.8   0.67     10.9
```

```
group_cc
```

```
## # A tibble: 10 x 5
## # Groups:   state_full [10]
##    state_full           rate_2001 rate_2016   diff diff_pct
##    <chr>                    <dbl>     <dbl>  <dbl>    <dbl>
##  1 RHODE ISLAND              9.45      5.99 -3.46    -36.6
##  2 NEW MEXICO                8.45      5.05 -3.4     -40.2
##  3 NORTH CAROLINA            8.5       5.75 -2.75    -32.4
##  4 NORTH DAKOTA              5.41      2.9  -2.51    -46.4
##  5 OHIO                      7.5       5    -2.5     -33.3
##  6 KANSAS                    6.49      4.67 -1.82    -28.0
##  7 MAINE                     8.75      7.15 -1.6     -18.3
##  8 UTAH                      6.01      5    -1.01    -16.8
##  9 OKLAHOMA                  5.91      5.07 -0.84    -14.2
## 10 DISTRICT OF COLUMBIA      9.38      8.84 -0.540    -5.76
```

```
mean(group_tt$diff)
```

```
## [1] 1.644
```

```
mean(group_tt$diff_pct)
```

```
## [1] 25.61677
```

```
mean(group_cc$diff)
```

```
## [1] -2.043
```

```
mean(group_cc$diff_pct)
```

```
## [1] -27.20366
```

```
tax_rate_merged %>%
  filter(state_full %in% group_tt$state_full) %>%
  filter(year == 2001 | year == 2016) %>%
  group_by(year) %>%
  select(state_full, year, millionaire_share) %>%
```

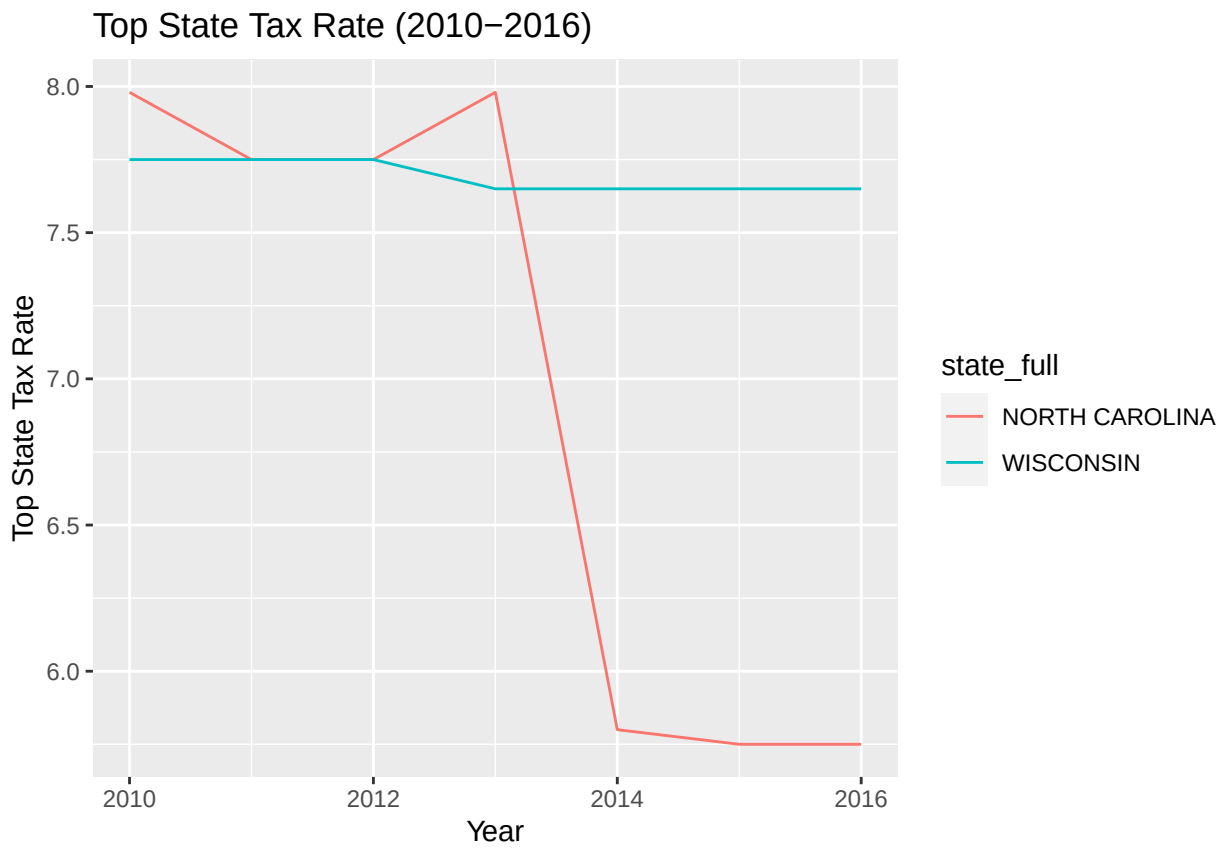
```
  pivot_wider(id_cols = state_full, names_from = year,
              values_from = c(millionaire_share)) %>%
  rename("mshare_2001" = 2) %>%
  rename("mshare_2016" = 3) %>%
  summarize(mshare_01 = sum(mshare_2001), mshare_16 = sum(mshare_2016)) %>%
  mutate(diff = mshare_16 - mshare_01, diff_pct = (mshare_16 - mshare_01) / mshare_01)
```

```
## # A tibble: 1 x 4
##   mshare_01 mshare_16    diff diff_pct
##       <dbl>     <dbl>   <dbl>    <dbl>
## 1     0.366     0.352 -0.0140  -0.0384
```

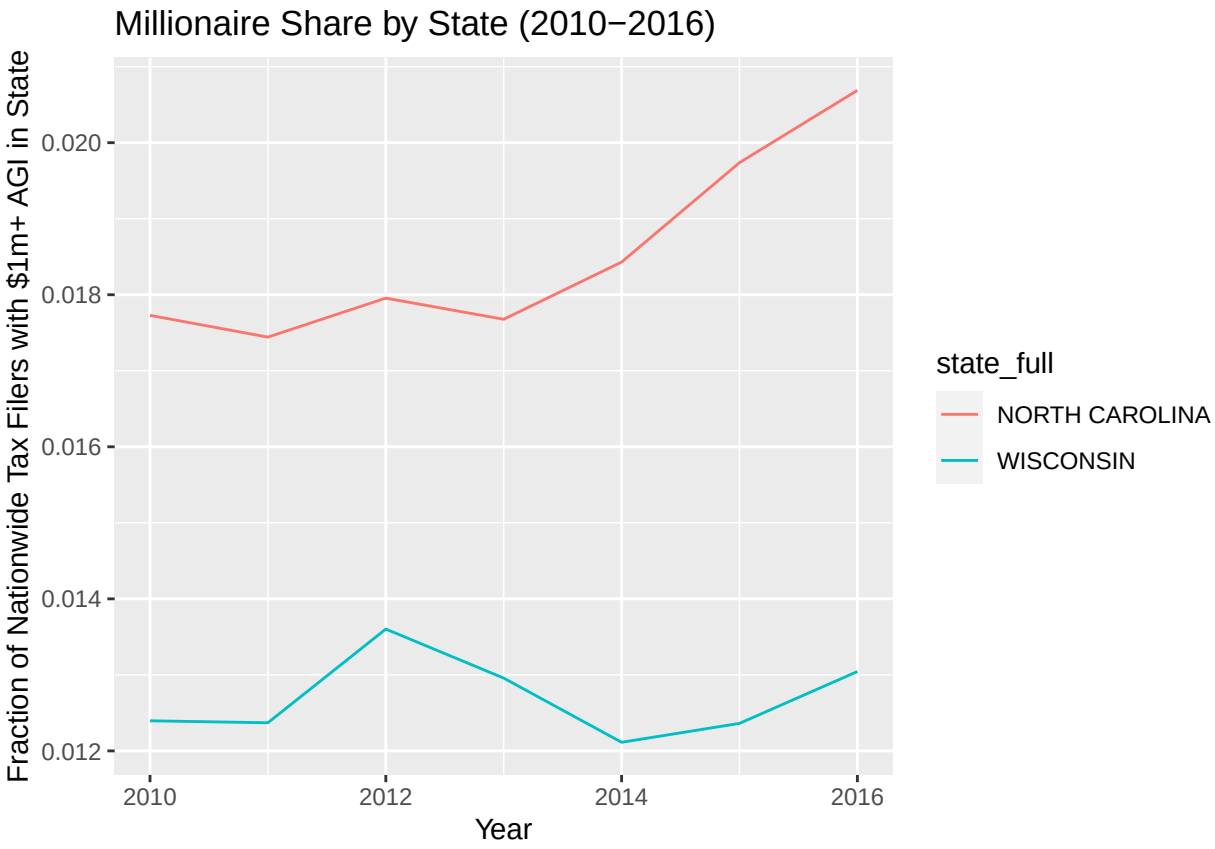
```
tax_rate_merged %>%
  filter(state_full %in% group_cc$state_full) %>%
  filter(year == 2001 | year == 2016) %>%
  group_by(year) %>%
  select(state_full, year, millionaire_share) %>%
  pivot_wider(id_cols = state_full, names_from = year,
              values_from = c(millionaire_share)) %>%
  rename("mshare_2001" = 2) %>%
  rename("mshare_2016" = 3) %>%
  summarize(mshare_01 = sum(mshare_2001), mshare_16 = sum(mshare_2016)) %>%
  mutate(diff = mshare_16 - mshare_01, diff_pct = (mshare_16 - mshare_01) / mshare_01)
```

```
## # A tibble: 1 x 4
##   mshare_01 mshare_16    diff diff_pct
##       <dbl>     <dbl>   <dbl>    <dbl>
## 1    0.0677    0.0771 0.00939    0.139
```

```
nc_wi <- tax_rate_merged %>%
  filter(state == "NC" | state == "WI") %>%
  filter(year <= 2016 & year >= 2010)

ggplot(data = nc_wi) + geom_line(aes(x = year, y = topstaterate, group = state_full,
                                     color = state_full)) +
  labs(title = "Top State Tax Rate (2010-2016)",
      x = "Year",
      y = "Top State Tax Rate")
```




```
ggplot(data = nc_wi) + geom_line(aes(x = year, y = millionaire_share, group = state_full,
                                     color = state_full)) +
  labs(title = "Millionaire Share by State (2010-2016)",
       x = "Year",
       y = "Fraction of Nationwide Tax Filers with $1m+ AGI in State")
```




```
question_1_6 <- nc_wi %>%
  filter(year != 2013) %>%
  mutate(control_years = case_when(
    year < 2013 ~ 1,
    year > 2013 ~ 0))

question_1_6 %>%
  group_by(state, control_years) %>%
  summarize(avg_million_share = mean(millionaire_share))
```

```
## `summarise()` has grouped output by 'state'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 4 x 3
## # Groups:   state [2]
##   state control_years avg_million_share
##   <chr>         <dbl>             <dbl>
## 1 NC                0            0.0196
## 2 NC                1            0.0177
## 3 WI                0            0.0125
## 4 WI                1            0.0128
```

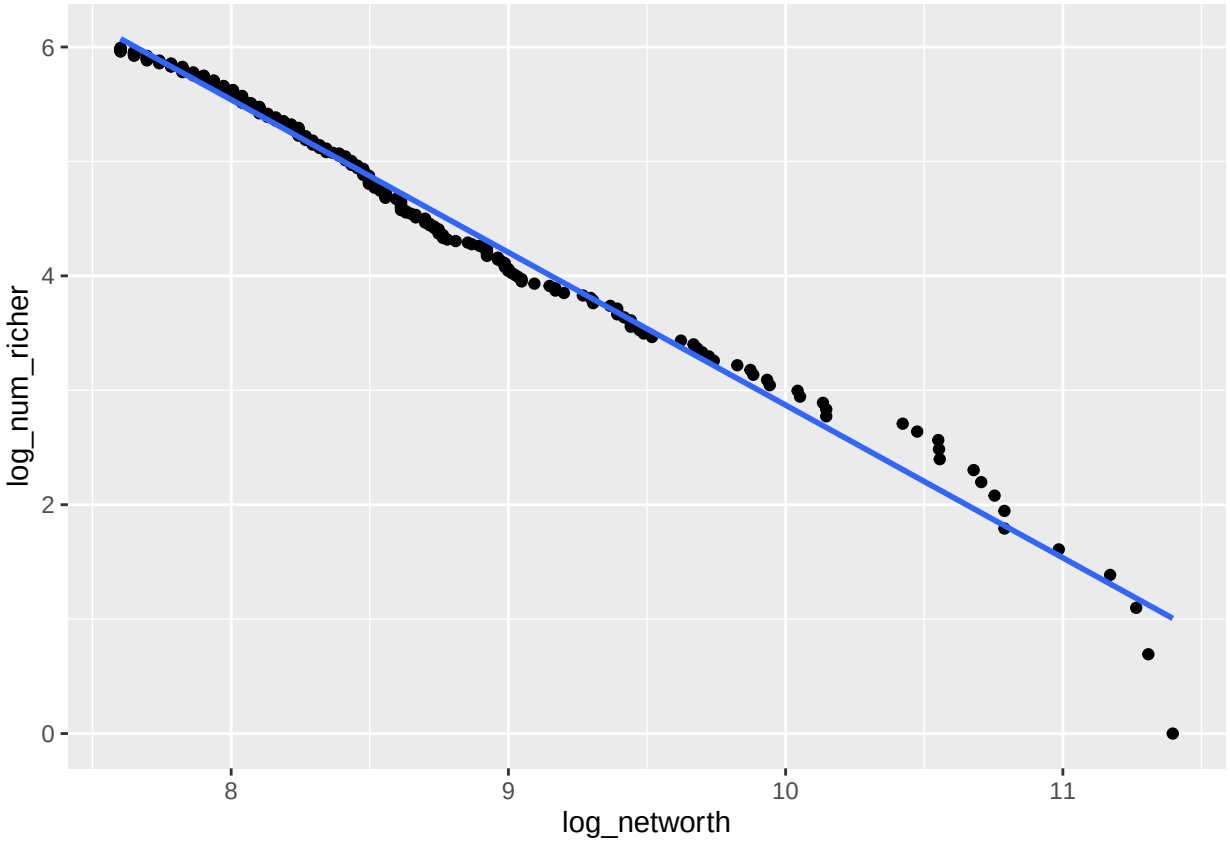
Question 2

The dataset forbes400.csv gives the list of the wealth of individuals in Forbes 400 in 2017.

1. Plot the number of individuals above a certain networth n as a function of networth n in a log-log scale. That is, plot log(Number of Individuals Above n) as a function of log(n).

```r
question_2_1 <- forbes400 %>%
  mutate(num_richer = 401 - row_number(),
         log_num_richer = log(num_richer),
         log_networth = log(networth))

ggplot(question_2_1, aes(x=log_networth, y=log_num_richer)) +
  geom_point() +
  geom_smooth(aes(x = log_networth, y = log_num_richer), method = "lm", se = F)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



2. As we discussed in class, the fact that this plot has linear slope in log-log space means that the wealth distribution is Pareto, and that the slope of the plot equals negative xi, where xi is the power law exponent of the distribution. Based on your plot, what is the power law exponent of the wealth distribution in 2017?

```r
lm(log_networth~log_num_richer, data=question_2_1)
```

```
##
## Call:
## lm(formula = log_networth ~ log_num_richer, data = question_2_1)
##
## Coefficients:
##    (Intercept)  log_num_richer
##        12.1083         -0.7408
```

3. What does it's magnitude imply for the magnitude of wealth inequality in the U.S. (hint: look at the

lecture notes for Lecture 9)

```
0.7408/(1 - 0.7408)
```

```
## [1] 2.858025
```

Beta = xi/(1 - xi)