p8105_hw4_kma2162_p2

Kathryn Addabbo

October 27, 2017

Load the appropriate libraries for this exercise.

```
library(janitor)
library(lubridate)
library(dplyr)
library(readr)
library(ggplot2)
```

Read and clean data:

The GDP dataset was read and the names were cleaned. The value column class was then mutated to a numeric, it contains the gross domestic production measurements. The column was then renamed gdp, to aid comprehension when merging with other datasets as the title value is generic. In this scenario, it is alright for NA's to be introduced via coercion because the first four rows do not contain a GDP value in the original dataset.

```
gdp = read_csv("./GDP.csv") %>%
  clean_names() %>%
  mutate(value = as.numeric(value)) %>%
  rename(gdp = value)
```

```
## Warning in evalq(as.numeric(value), <environment>): NAs introduced by
## coercion
```

The pols-month dataset was read and names were cleaned. The mon column class was mutated to date with the appropriate format to recognize that year came first, then month, then day. The lubridate function was used to round down each day to the first of the month and a new column date was created with those values. Therefore the mon column was removed and the prez_gop column, as it is a replication of the prez_dem column (with the value of 0 and 1 reversed). Then the column was given a more generic name of prez.

```
pols_month = read_csv("./pols-month.csv") %>%
  clean_names() %>%
  mutate(mon = as.Date(mon, format = "%Y-%m-%d")) %>%
  mutate(date = lubridate::floor_date(mon, "month")) %>%
  select(-c(mon, prez_gop)) %>%
  rename(prez = prez_dem)
```

To make comprehension of this column more clear, the values inside were changed to read dem and gop when the president was democratic and republican respectively.

```
pols_month$prez[pols_month$prez == "1"] = "dem"
pols_month$prez[pols_month$prez == "0"] = "gop"
```

The two datasets can then be merged by date.

```
pol_data = merge(gdp, pols_month, by = "date")
```

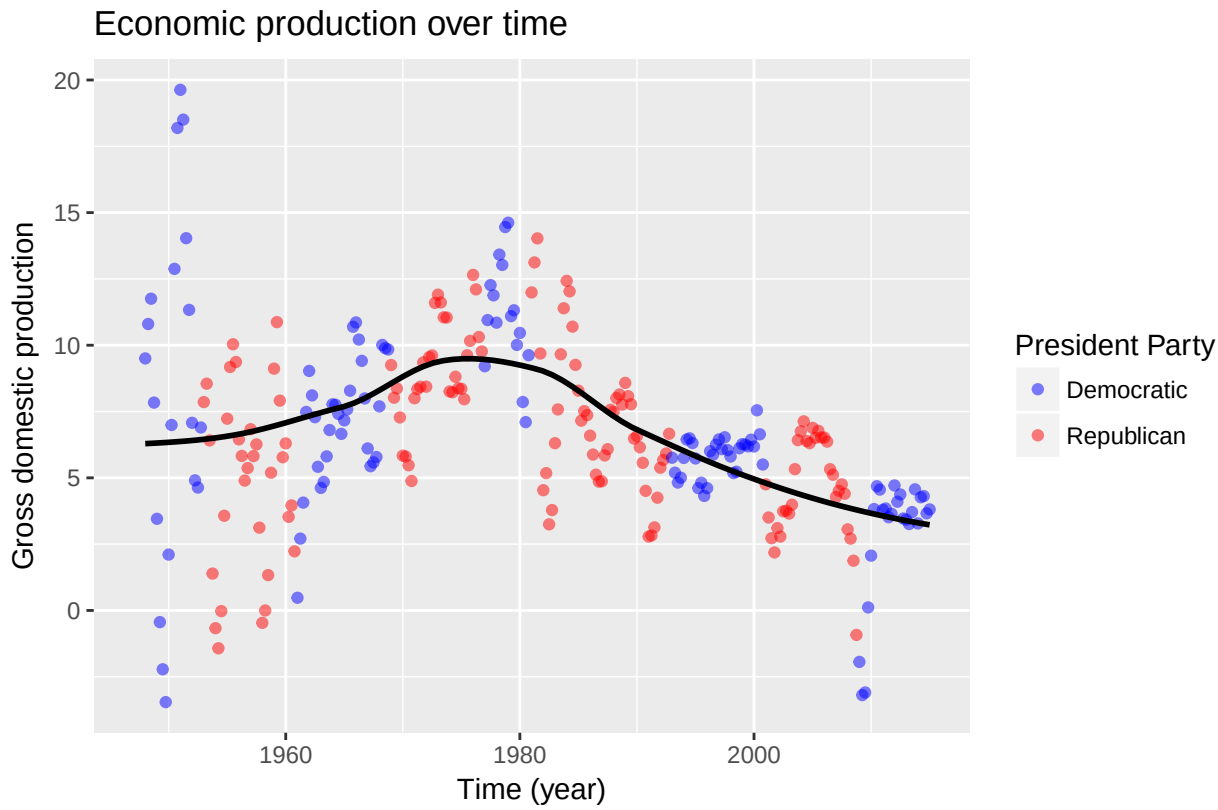
The following plot showcases the evaluation of GDP over time, while taking into account the party of the president.

```r
ggplot(pol_data, aes(x = date, y = gdp)) +
  geom_point(aes(color = prez), alpha = 0.5) +
  scale_color_manual(labels = c("Democratic", "Republican"),
                           values = c("blue", "red")) +
  guides(color = guide_legend("President Party")) +
  geom_smooth(se = FALSE, color = "black") +
  labs(
     title = "Economic production over time",
     x = "Time (year)",
     y = "Gross domestic production",
     caption = "Data showing the fluctation of economic growth while taking into account the president's
   )
```

```
## `geom_smooth()` using method = 'loess'
```

```
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```



tion of economic growth while taking into account the president's political affiliation.

Observing the plot, we notice that economic growth is increasing and peaks right before 1980. After that, the US experiences a steady decline in which we know there to be four recessions over the course of said 30 years. However, there does not seem to be a drastic change in production when observing presidential affiliation. It can be noted that as time progresses, the range of GDP recorded per presidency decreases.