Assignment from Nov 19

Hasaan Parker

11/19/2020

Create the following graphs in `ggplot2`.

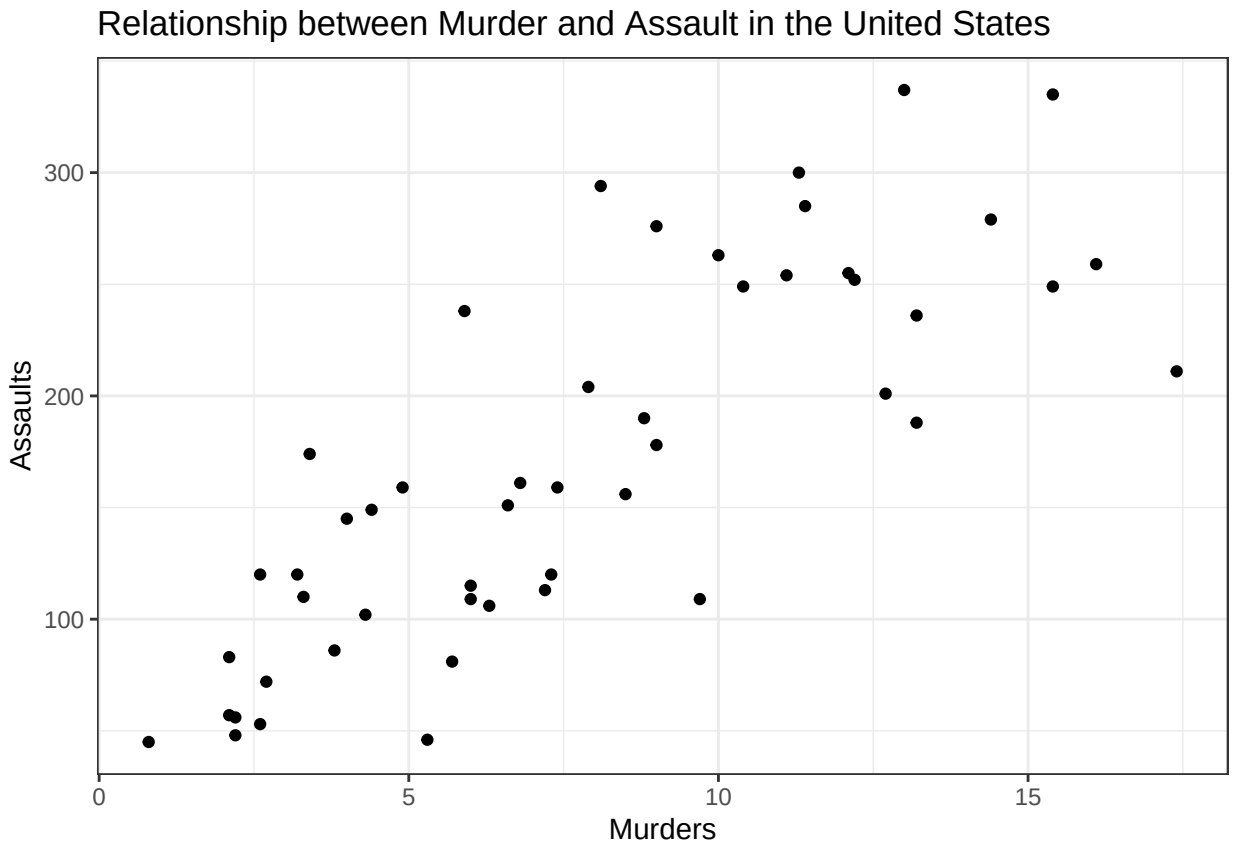
1. Check out the base R built-in dataset, `data("USArrests")`.

```
data("USArrests")
summary(USArrests)
```

```
##       Murder          Assault         UrbanPop          Rape
##  Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30
##  1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07
##  Median : 7.250   Median :159.0   Median :66.00   Median :20.10
##  Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23
##  3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18
##  Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00
```

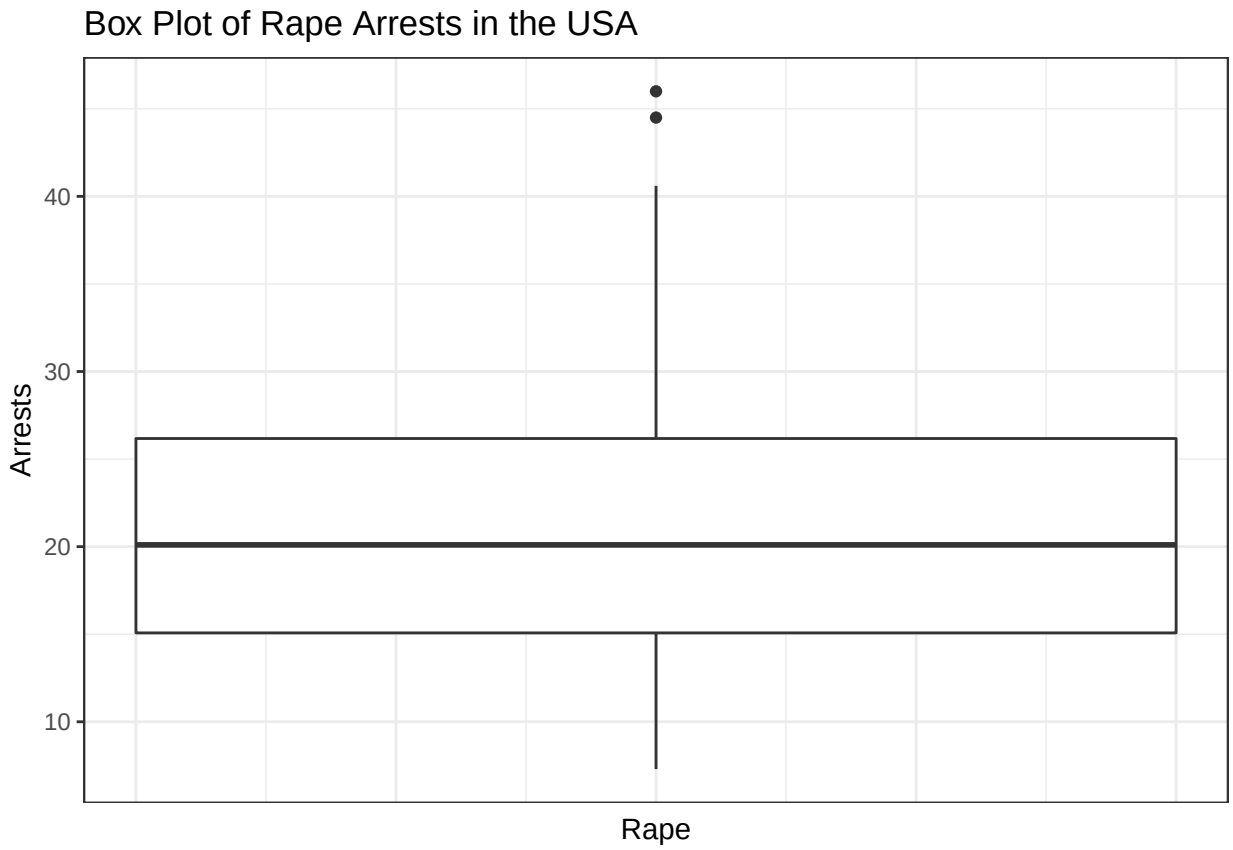
2. Create a scatterplot that looks at the correlation between murder and assault arrests. Label the x and y axes and title the graph.

```
USArrests %>% ggplot(aes(Murder, Assault)) +
  geom_point() +
  labs(title = "Relationship between Murder and Assault in the United States",
       x = "Murders", y = "Assaults")
```



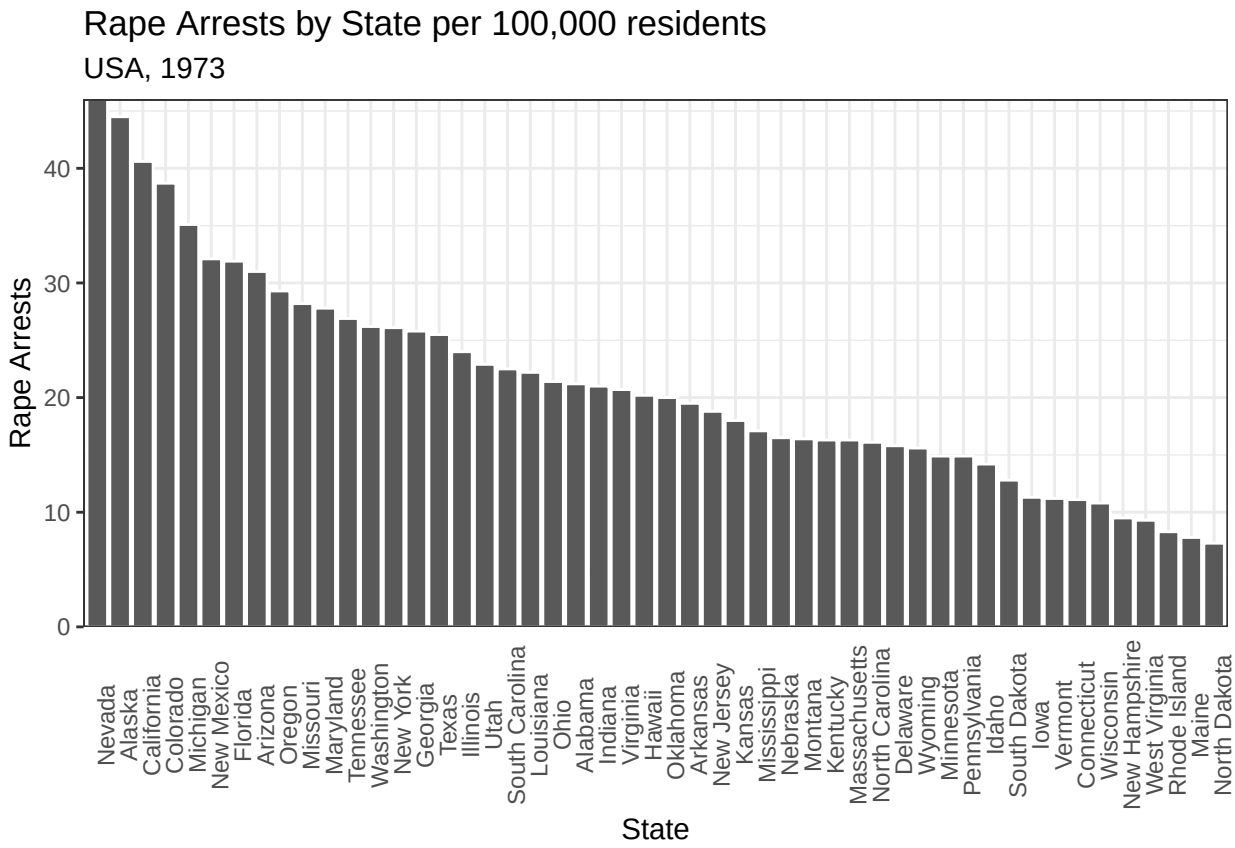
3. Create a boxplot of rape arrests. Label the plot.

```
USArrests %>% ggplot(aes(y = Rape )) +
  geom_boxplot(width = .1) +
  labs(title = "Box Plot of Rape Arrests in the USA",
       x = "Rape", y = "Arrests") +
  theme(axis.ticks.x = element_blank(),
        axis.text.x = element_blank())
```



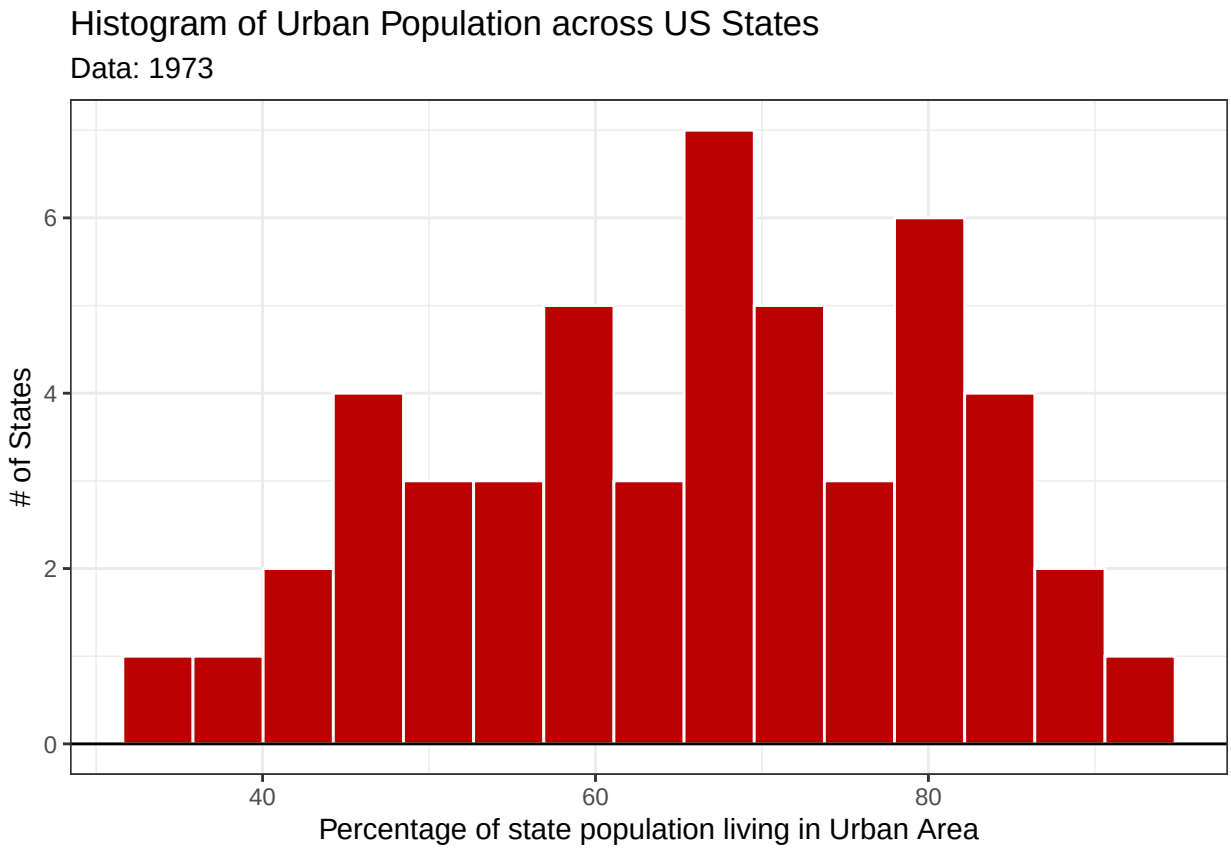
4. Create a barplot of the number of rape arrests per state.

```
USArrests %>%
  arrange(desc(Rape)) %>%
  rownames_to_column("State") %>%
  ggplot(aes(reorder(State, -Rape), Rape)) +
  geom_col(color = "white") +
  scale_y_continuous(expand = c(0,0)) +
  labs(title = "Rape Arrests by State per 100,000 residents",
       subtitle = "USA, 1973 ",
        x = "State", y = "Rape Arrests") +
  theme(axis.text.x = element_text(angle=90),
        axis.ticks.x = element_blank())
```



5. Create a histogram for the percent of urban population.

```
USArrests %>%
  ggplot(aes(UrbanPop)) +
  geom_histogram(color = "white", fill = "#BB0000", bins = 15) +
  geom_hline(yintercept = 0) +
  labs(title = "Histogram of Urban Population across US States",
       subtitle = "Data: 1973",
       x = "Percentage of state population living in Urban Area",
       y = "# of States"
       )
```



Your project

Now it's your turn. Use the `ggplot2` tools you used today to conduct data analysis for one of your final seminar papers.

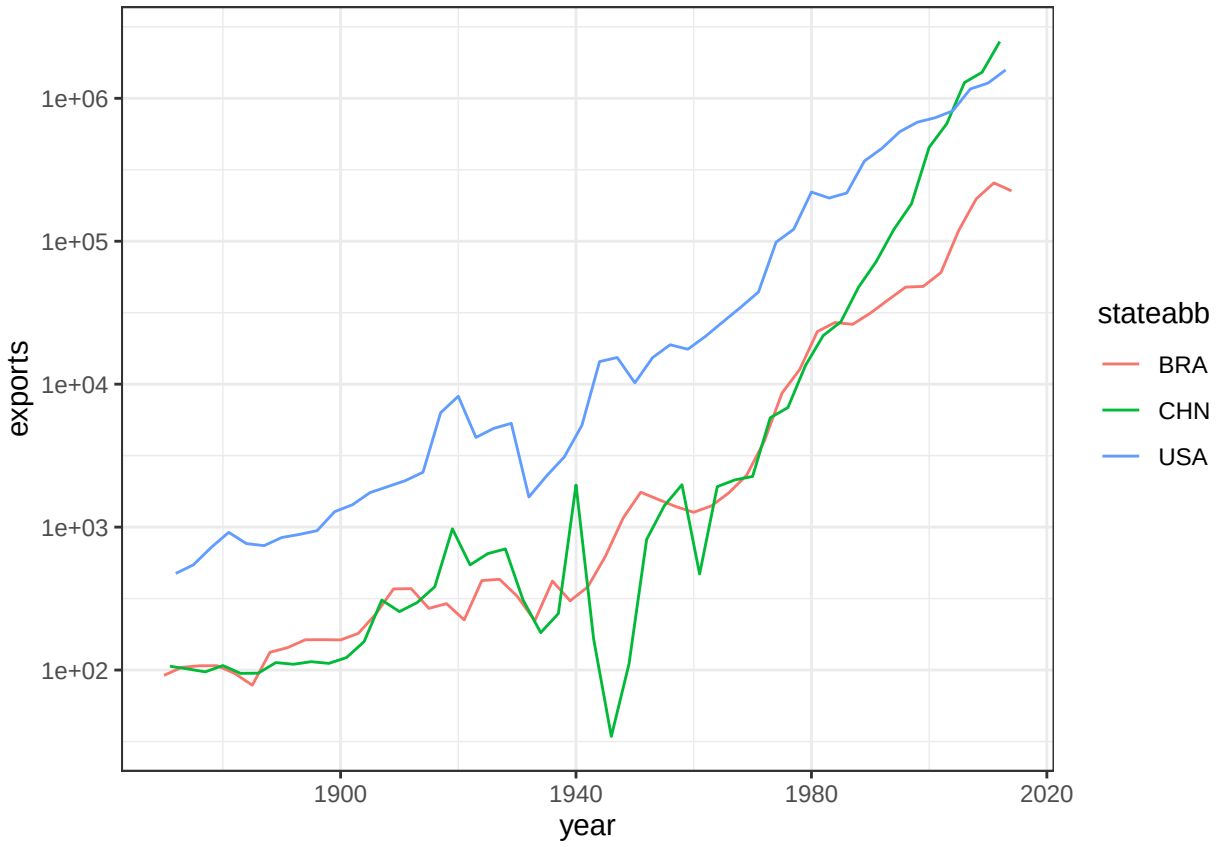
1. Create a Git repository for your project.
2. Upload the dataset you are planning to use. If are you planning to collect original data, please provide all the variables that will be in that original dataseet.
3. How will these variables help answer your question? I will be using the CoW Project national trade data to analyze patterns in international trade overtime.
4. Consider the variables. Which variables would you want to highlight? How would you visually represent them? Plan to create at least three descriptive graphs. Some suggestions:
    - If you have time-related variable, create a line graph showing changes over time!
    - If your observations can be separated by certain groups, create bar graphs or facets~ I plan to create a line graph of international export flow over for the US, China, and Brazil. Then, I will create a bar graph of the import/export balance for the largest 25 countries in the most recent year of the dataset (2014). Lastly, I will make a graph comparing the import/export balance to total trade volume, to understand if there is a relationship between the size of an economy and whether it imports or exports more goods/services.
5. Given what you brainstormed in Question 4, create plots using the `ggplot2` package. Label all axes and title each graph. Provide descriptions for each graph.

```
trade <- read_csv("/Users/hasaanparker/Documents/School/Fall 21/812 stats intro/Final Project/Data/COW_
```

```
## Rows: 14410 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): statename, stateabb
## dbl (9): ccode, year, imports, exports, alt_imports, alt_exports, source1, s...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

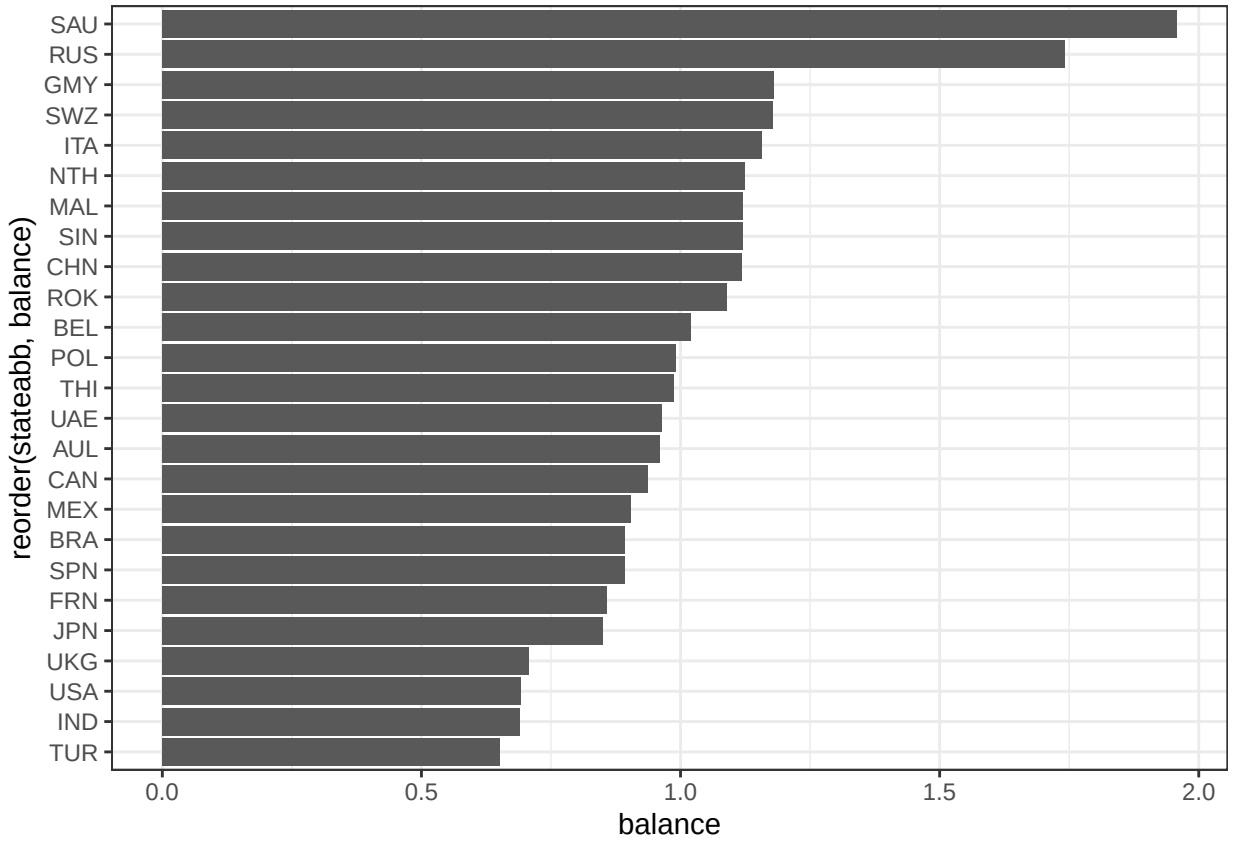
```
trade %>%
  filter(stateabb == c("CHN", "BRA", "USA")) %>%
  ggplot(aes(year, exports))  +
  geom_line(aes(col = stateabb)) +
  scale_y_log10()
```

```
## Warning in stateabb == c("CHN", "BRA", "USA"): longer object length is not a
## multiple of shorter object length
```




```
trade <- trade %>%
  mutate(total = imports + exports,
         balance = exports / imports)
trade %>%
  filter(year == 2014) %>%
  arrange(desc(total)) %>%
  slice_head(n = 25) %>%
  ggplot(aes(reorder(stateabb, balance), balance)) +
  geom_col() +
```

```
  coord_flip()
```

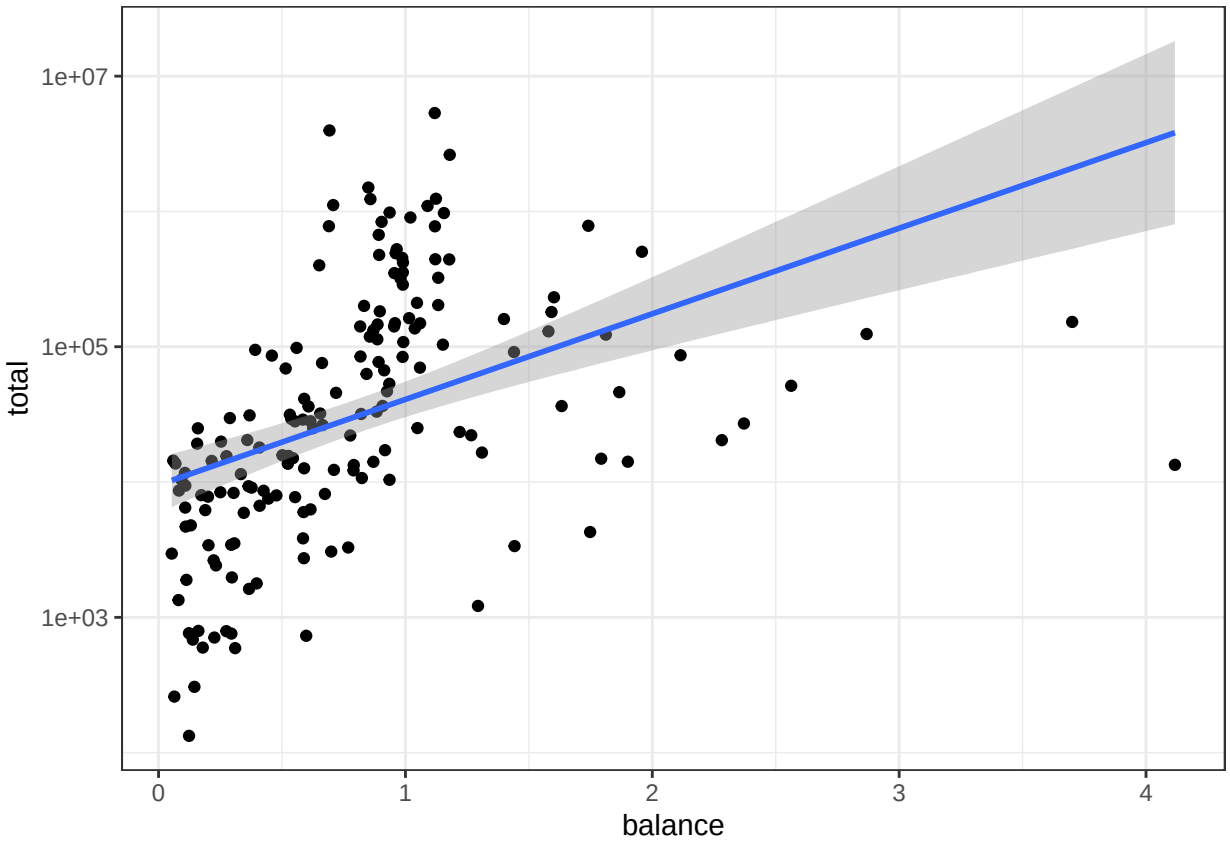



```
trade %>%
  filter(year == 2014) %>%
  ggplot(aes(balance, total)) +
  geom_point() +
  geom_smooth(method = "lm") +
  scale_y_log10()
```

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

```
## Warning: Removed 9 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```



Submit

Email me (laaker@wisc.edu) the link to your github when you are done.