PS2_answers

vkvats

1/23/2020

Q1.

Importing data

```
library(MASS)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------------------
## v tibble  2.1.3      v dplyr   0.8.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
## v purrr   0.3.3
## -- Conflicts -----------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()
```

```
# Histrogram plot
ggplot(cats, aes(x = Bwt)) + geom_histogram(binwidth = 0.1) +
  ggtitle("Cat weight from Fisher (1947)") +
  xlab("Bodyweight (kg)") +
  ylab("Number of cats")
```

Cat weight from Fisher (1947)

Description of the distribution:

Looking at the histogram plot of the cats weight, it can be said that the distribution seems to be right skewed as the plot extends more towards right of the plot. the weight of the cats varies from 2 kg to 3.9 kg with mean value of the distribution is 2.72 kg and median of the distribution is 2.7 kg. The QQ plot is fairly straight (not shown here) apart from the end points where it doens't really follow the straight line. since the data is right skewed, we can apply log transformation if we wish to do conformatory data analysis.

Q2. ANES2016

```
ANES2016.df <- data.frame(read.csv("anes_pilot_2016.csv"))
thermometer <- ANES2016.df[,c("fttrump", "fthrc", "ftsanders", "ftrubio")]
names(thermometer) <- c("Trump", "Hillary", "Sanders", "Rubio")

feeling.df <- stack(thermometer)
names(feeling.df)<- c("thermo_feeling", "leader")
response.df <- feeling.df[feeling.df$thermo_feeling <= 100,]
```

Plot: Draw ONE graph that clearly shows the differences in the shape of the distributions of feeling thermometer scores for Clinton, Sanders, Rubio, and Trump.

```
ggplot(response.df, aes(x = thermo_feeling)) + geom_histogram(binwidth = 1) +
  facet_wrap(~leader) +
  ggtitle("Feeling thermometer for each candidate") +
  ylab("Count") +
  xlab("Feeling thermometer score")
```

Feeling thermometer for each candidate
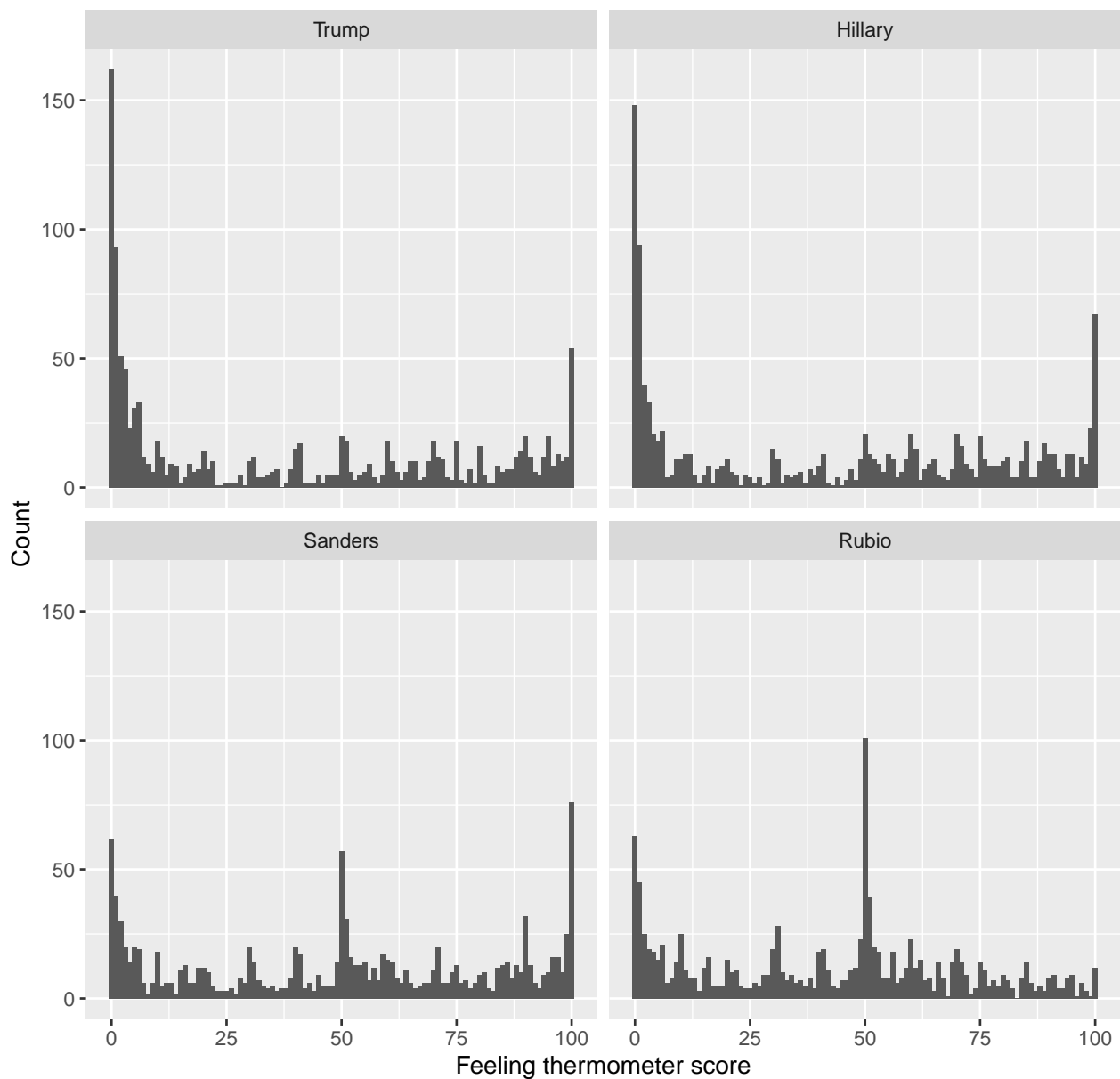

```r
# calculating summary of thermo_feeling for all leaders
central.values <- response.df %>%
  group_by(leader) %>%
summarise(Mean = mean(thermo_feeling), Median = median(thermo_feeling), SD = sd(thermo_feeling))
central.values
```

```
## # A tibble: 4 x 4
##   leader   Mean Median    SD
##   <fct>   <dbl>  <dbl> <dbl>
## 1 Trump    38.4     30  36.5
## 2 Hillary  43.0     44  36.5
## 3 Sanders  50.4     51  33.4
## 4 Rubio    41.5     47  28.1
```

Description of graph:

(a) **Trump** : The mean value of the thermometer feeling is approximately 38, which is above its median value 30. There are more people who has very cold feeling for Trump. This is also evident from the histogram plot, where there are 581 people out of 1197 (valid response) who has given cold feeling of less than or equal to 25 and there are 283 people who has given warm feeling in between 75 to 100 for Trump. The most occuring thermometer feeling for Trump at 'zero', almost 162 people have given him 'zero' thermometer feeling.

(b) **Hillary** :The mean value of the thermometer feeling is approximately 43 , which is close to its median value 44. The distribution is more concentrated at the beginning, which shows the peak value at zero (around 148 people has reported this). Around 502 people out of 1199 (valid responses) who has given cold feeling of less than or equal to 25 and there are 326 people who has given warm feeling in between 75 to 100.

(c) **Sanders** : The mean value of the thermometer feeling is approximately 50 , which is close to its median value 51. The distribution is there major areas of concentration around zero, around 50 and around 100. Around 340 people out of 1192 (valid responses) who has given cold feeling of less than or equal to 25 and there are 349 people who has given warm feeling in between 75 to 100 for Trump. The peak of the distribution is around 100 where 76 (approx) poeple has shown warm feeling for him.

(d) **Rubio** : The mean value of the thermometer feeling is approximately 41 , which is close to its median value 47. This indicates that the distribution is right skewed and there are very few people who has warm feeling for Rubio (166 people who rating between 75 and 100 out of 1187 valid responses), which is evident from plot as well. The distribution peaks at 50 with 101 people. Around 381 people have given cold feeling of less than or equal to 25.

```r
# replicating the immig_numb four time to add the column of immig_number for each leader.
immigration <-  data.frame(rep(ANES2016.df$immig_numb, 4))
immig_feeling.df <- data.frame(feeling.df, immigration)
names(immig_feeling.df)<- c("thermo_feeling", "leader", "immig_numb")
new.immig.feeling.df <- immig_feeling.df[c(which(immig_feeling.df$thermo_feeling <= 100)),]

# calculating the mean value of thermomter feeling for each level of immigration for each leader.
mean.values = new.immig.feeling.df %>%
  group_by(leader, immig_numb) %>%
  summarise(Mean = mean(thermo_feeling))
```

Plot:(b) Draw ONE graph that clearly shows the differences in the mean feeling thermometer score for each level of immigration attitude for each of the four candidates.

```r
#ggplot(new.immig.feeling.df, aes(y = thermo_feeling)) + geom_boxplot() + facet_wrap(~immig_numb)
ggplot(mean.values, aes(x = immig_numb, y = Mean)) + geom_point() +
  facet_wrap(~leader, ncol = 2) +
  labs(title = "Mean feeling trend for immigration attitudes",
       subtitle = "For each level of immigration attitude for each candidates") +
  xlab("Levels of immigration attitude") +
  ylab("Feeling thermometer")
```

Mean feeling trend for immigration attitudes

For each level of immigration attitude for each candidates
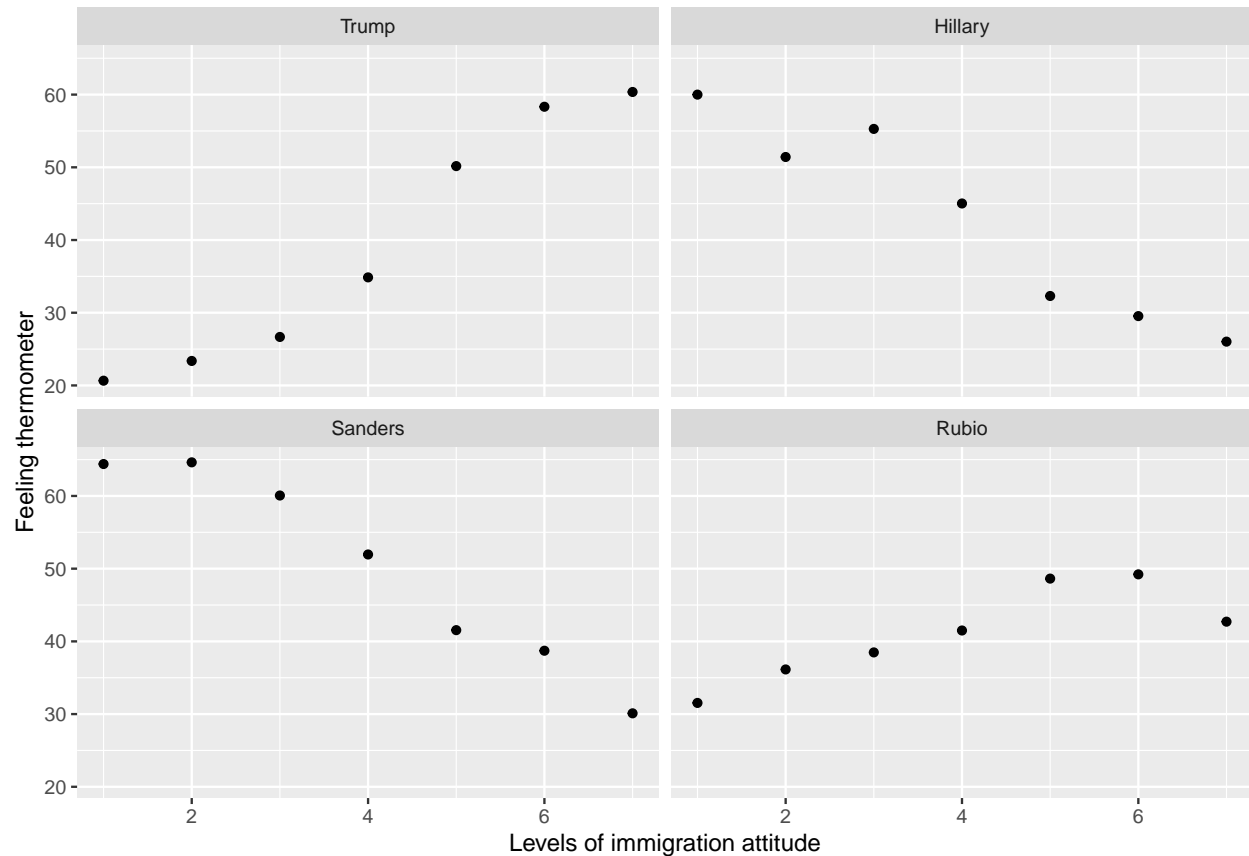
Description: It can be observed from the plot that the mean value of thermometer feeling for Trump increases from immigration attitude level 1 - increased a lot, 20.65, to level 7 - decreased a lot, 60.36, this trend is also same for Rubio, goes from 31.53, for level 1, to 49.21 for level 6 then it drecreses for level 7 to 42.7, but the increase in mean values are lesser than Trump's level with overall trend going upward. For Hillary, the mean thermometer feeling trends downward from level 1 (feeling value of 60) to level 7 (feeling value 26). The peak values of Trump at level 7 is close to peak value of Hillary at level 1, which shows the contrast in between them. Sanders follows a similar trend as Hillary, from level 1 score of 64 to down to score 30 at level 7. The over all trend is same for Hillary and Sanders and also the overall trend is same for Trump and Rubio.