EDA on Election data

2023-09-07

```
data<- read_csv("Assignment 2 data.csv", show_col_types = FALSE)
view(data)
```

#PART 1 Removing missing values

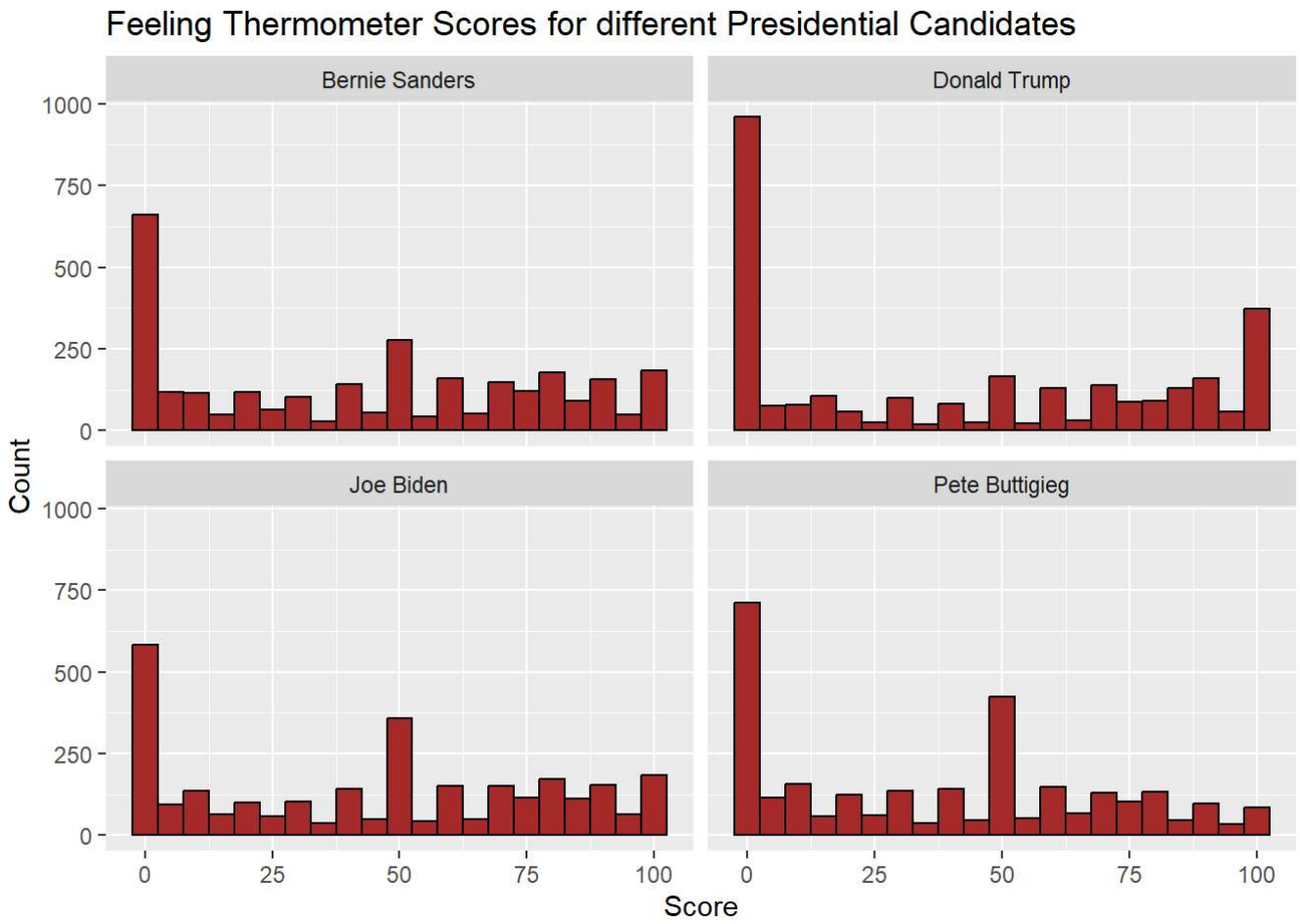
```
data<-data[data$ftbiden1<=100 & data$fttrump1<=100 & data$ftsanders1<= 100 & data$ftbuttigieg1<=
100,]
```

Creating a new data frame

```
df<-data.frame(
  Candidate = rep(c("Joe Biden", "Donald Trump", "Bernie Sanders", "Pete Buttigieg"), each = nro
w(data)),
  Score = c(data$ftbiden1, data$fttrump1, data$ftsanders1, data$ftbuttigieg1))
view(df)
```

Plotting

```
ggplot(df, aes(x=Score))+
  geom_histogram(color="black",fill="brown", binwidth=5)+
  labs(title="Feeling Thermometer Scores for different Presidential Candidates",
       x="Score",
       y="Count")+
  facet_wrap(~Candidate)
```



#Conclusion -Donald Trump received the least feeling thermometer score of 0-5 the most ;around 900 people gave Trump least score whereas the other three candidates received less than 750 votes with least score.

-Also Trump received 100 score from around 375 people whereas others received less votes on 100. So it can be said that Trump has both haters and lovers(more haters than lovers).

-Most common feeling thermometer score received are 0, 50 and 100 except for Pete Buttigieg. Pete Buttigieg received the 100 score the least times.

-Bernie Sanders and Joe Biden have similar distribution for feeling thermometer score, which means people feel kind of same way for both of them.

#PART 2 Dropping row having missing value

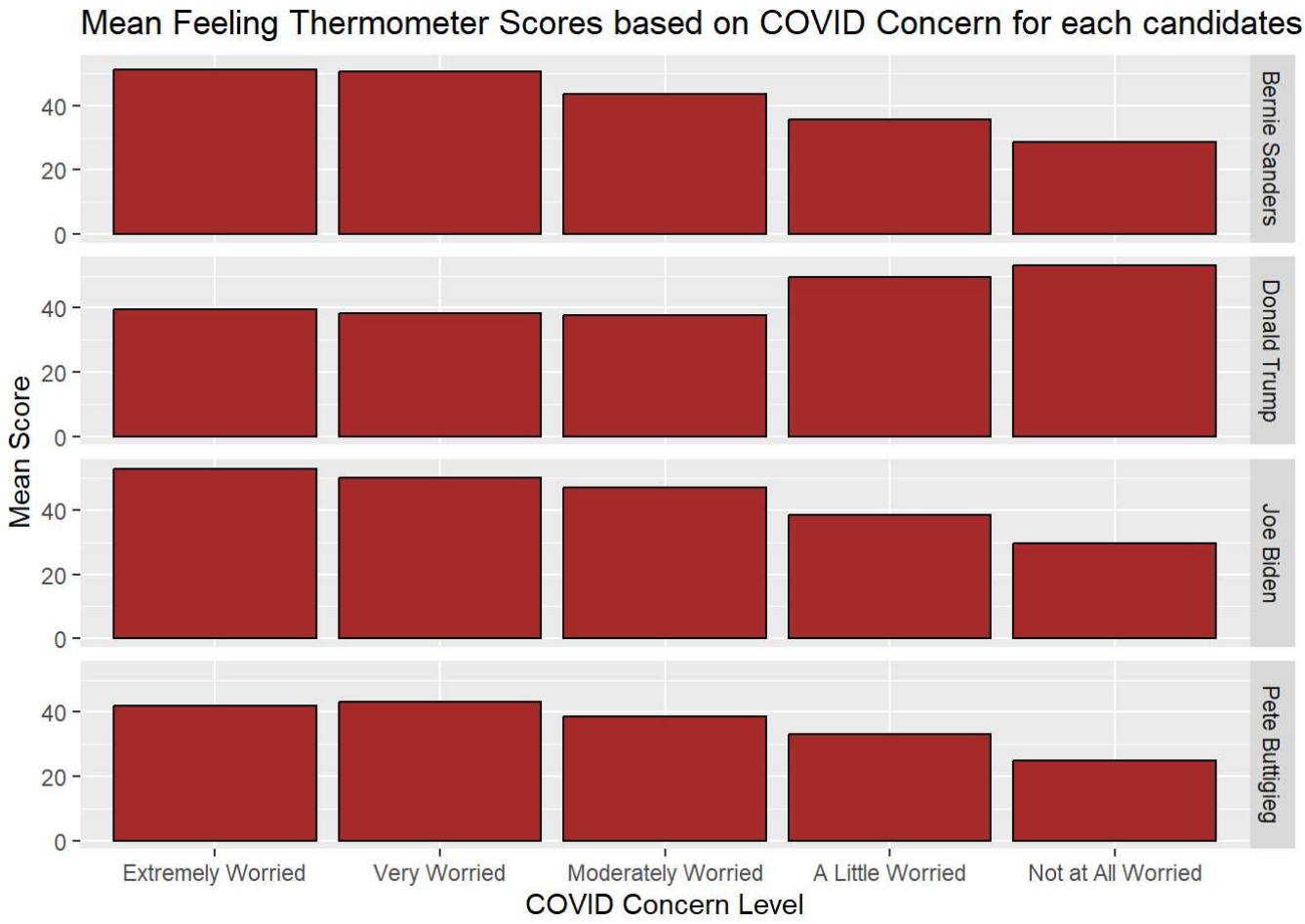
```
data<- data[data$covid1 != 9,]
```

Creating a new data frame)

```
df2<- data.frame(
  COVID_Concern=factor(data$covid1, levels=1:5, labels= c("Extremely Worried", "Very Worried",
"Moderately Worried", "A Little Worried", "Not at All Worried")),
  Candidate=rep(c("Joe Biden", "Donald Trump", "Bernie Sanders", "Pete Buttigieg"), each=nrow(da
ta)),
  Score=c(data$ftbiden1, data$fttrump1, data$ftsanders1, data$ftbuttigieg1))
view(df2)
```

```
mean_score <- aggregate(Score ~ COVID_Concern + Candidate, data = df2, FUN = mean)
view(mean_score)
```

```
ggplot(mean_score, aes(x = COVID_Concern, y= Score)) +
  geom_bar(stat = "identity", color="black", fill="brown") +
  labs(title = "Mean Feeling Thermometer Scores based on COVID Concern for each candidates",
       x = "COVID Concern Level",
       y = "Mean Score")+
  facet_grid(Candidate~.)
```



#Conclusion -The overall distribution of feeling thermometer score with respect to Covid concern for the presidential candidates is similar except for Trump.

-People who are mostly concerned about Covid have given the maximum average score to almost all candidates.

-Joe Biden has the maximum mean feeling thermometer score of little greater than 50 from the extremely worried category of people. Just behind him is Bernie Sanders with around 50.

-Different from other cases Donald Trump received the maximum mean score from the ‘not at all worried’ category of people.

-For all candidates except Trump, people who are most concerned about Covid gave maximum score and vice versa. Trump received most scores from comparitively less concerned people.