# EDA_Assignment2

Ojaas Hampiholi

1/27/2020

# Question 1

```
library(MASS)
library(ggplot2)
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.6.2
```

```
mean(cats$Bwt)
```

```
## [1] 2.723611
```

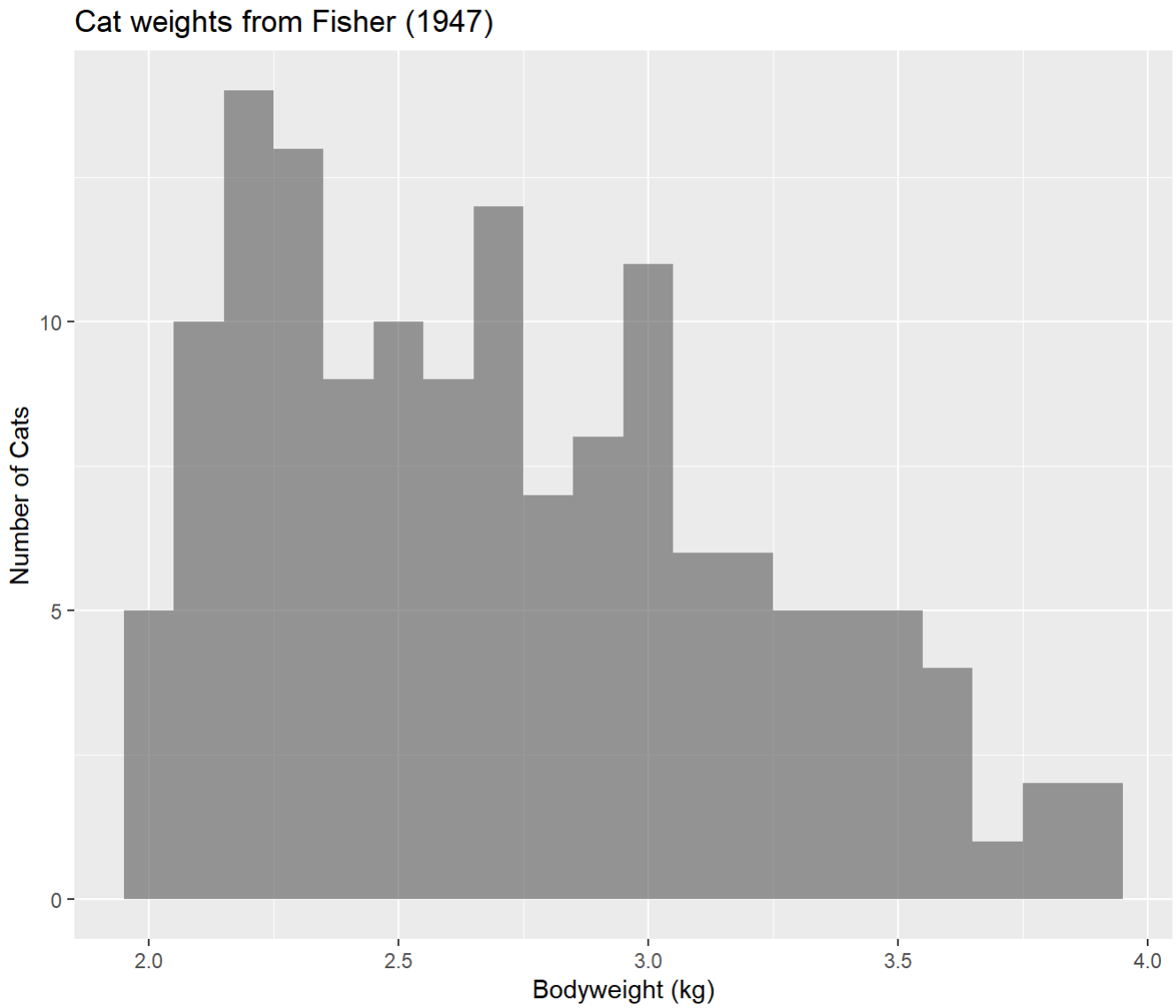
```
var(cats$Bwt)
```

```
## [1] 0.2355225
```

```
median(cats$Bwt)
```

```
## [1] 2.7
```

```
ggplot(cats, aes(x = Bwt)) + geom_histogram(binwidth = 0.1, alpha = 0.6) +
  xlab("Bodyweight (kg)") + ylab("Number of Cats") + ggtitle("Cat weights from Fisher (1947)")
```




```
#ggplot(cats, aes(x = Bwt, y = Hwt)) + geom_point() + xlab("Bodyweight (kg)") + ylab("Number of
 Cats")+ ggtitle("Cat weights from Fisher (1947)")
```

The plot above clearly shows that the given distribution is a right skewed distribution. The weights of cats are concentrated more between 2 and 3 kgs, while the value higher than it are rare occurances. The numeric data shows that the mean weight of cats is 2.72 kgs with a variance of 0.235 kgs. The median value of the cats weights in the data set is 2.7

# Question 2

## Subpart A

```
ANESraw = read.csv("C:/EDA/Assignment 2/anes_pilot_2016.csv", header = TRUE)
#View(ANESraw)
```

```
newData = subset(ANESraw, fttrump <= 100 & fthrc<=100 & ftsanders <= 100 & ftrubio <= 100)
presidentialCandidates = gather(newData, "FeelingThermometer", "Response", c("fttrump", "fthrc",
"ftsanders", "ftrubio"))
mean(newData$fttrump)
```

```
## [1] 38.27203
```

```
#sd(newData$fttrump)
mean(newData$fthrc)
```

```
## [1] 43.10763
```

```
#sd(newData$fthrc)
mean(newData$ftsanders)
```

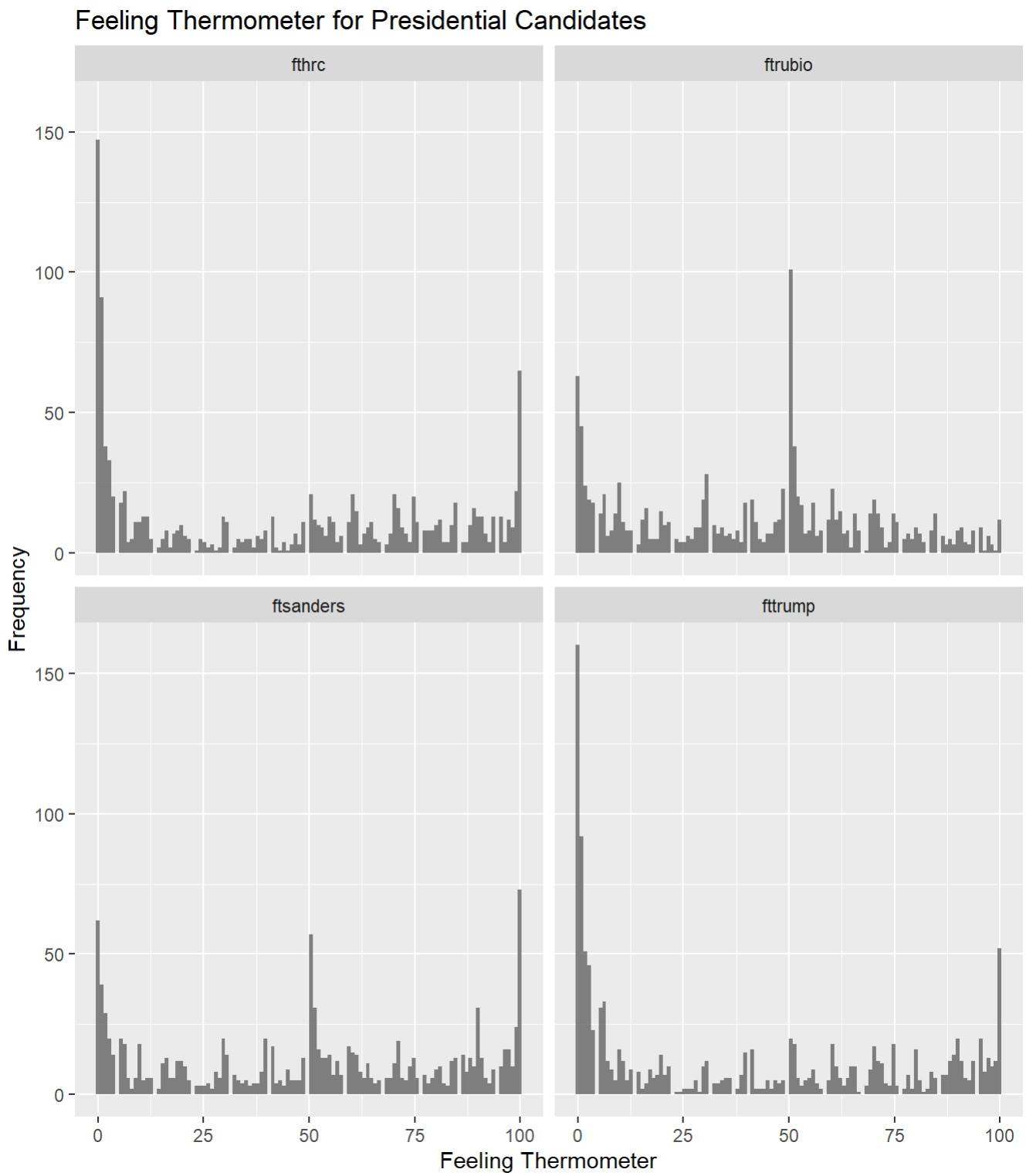
```
## [1] 50.30254
```

```
#sd(newData$ftsanders)
mean(newData$ftrubio)
```

```
## [1] 41.51356
```

```
#sd(newData$ftrubio)

ggplot(presidentialCandidates, aes(x = Response)) + geom_histogram(binwidth = 0.9, alpha = 0.75)
+
  facet_wrap(~ FeelingThermometer, ncol = 2) + xlab("Feeling Thermometer") + ylab("Frequency")+
 ggtitle("Feeling Thermometer for Presidential Candidates")
```




From the data summary and the graphs plotted above, we can reach a conclusion that: 1. There a lot of people who have cold feelings towards Donald Trump, the data shows that the mean reading for Donald Trump on the Feeling Thermometer is 38.27 2. There a lot of people who have cold feelings towards Hillary Clinton, the data shows that the mean reading for Hillary Clinton on the Feeling Thermometer is 43.10 3. Bernie Sanders has a few cold feelings and a few hot feelings on the feeling thermometer, which means that overall he has the highest support on this scale. This can be proved by mean for the data, which is placed at 50.30 4. A lot people have mixed feelings about Marco Rubio. The mean for his data is placed at 41.51

# Subpart B

```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
newData$immig_numb = factor(newData$immig_numb, order = TRUE)
#summary(newData)

trump_1 = mean(ANESraw[ANESraw$immig_numb == 1, ]$fttrump)
trump_2 = mean(ANESraw[ANESraw$immig_numb == 2, ]$fttrump)
trump_3 = mean(ANESraw[ANESraw$immig_numb == 3, ]$fttrump)
trump_4 = mean(ANESraw[ANESraw$immig_numb == 4, ]$fttrump)
trump_5 = mean(ANESraw[ANESraw$immig_numb == 5, ]$fttrump)
trump_6 = mean(ANESraw[ANESraw$immig_numb == 6, ]$fttrump)
trump_7 = mean(ANESraw[ANESraw$immig_numb == 7, ]$fttrump)
trump = c(trump_1, trump_2, trump_3, trump_4, trump_5, trump_6, trump_7)
summary(trump)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.37   29.48   36.98   42.24   54.25   67.86
```

```
hillary_1 = mean(ANESraw[ANESraw$immig_numb == 1, ]$fthrc)
hillary_2 = mean(ANESraw[ANESraw$immig_numb == 2, ]$fthrc)
hillary_3 = mean(ANESraw[ANESraw$immig_numb == 3, ]$fthrc)
hillary_4 = mean(ANESraw[ANESraw$immig_numb == 4, ]$fthrc)
hillary_5 = mean(ANESraw[ANESraw$immig_numb == 5, ]$fthrc)
hillary_6 = mean(ANESraw[ANESraw$immig_numb == 6, ]$fthrc)
hillary_7 = mean(ANESraw[ANESraw$immig_numb == 7, ]$fthrc)
hillary = c(hillary_1, hillary_2, hillary_3, hillary_4, hillary_5, hillary_6, hillary_7)
summary(hillary)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   26.02   30.92   45.03   43.84   56.99   60.00
```

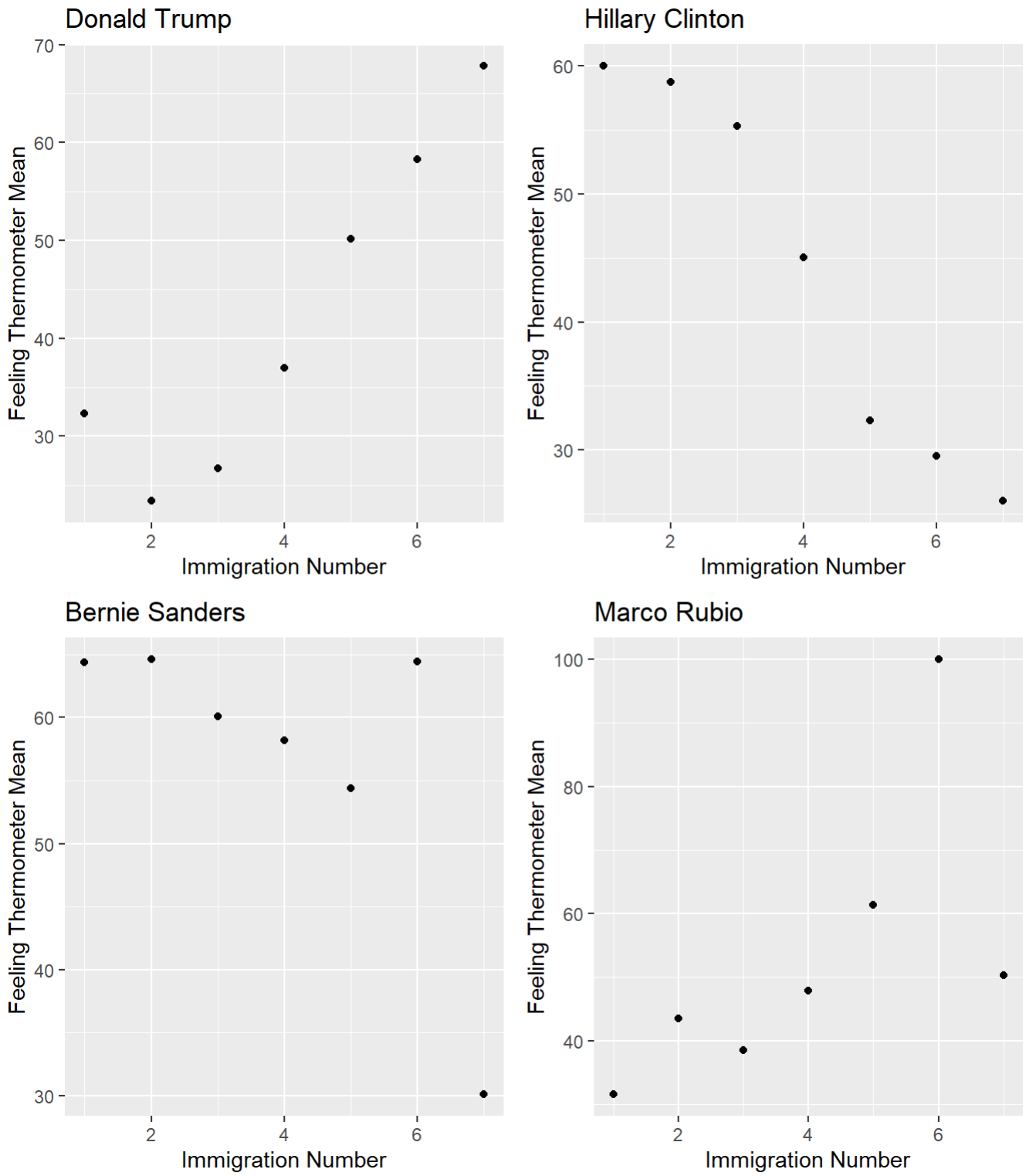
```
sanders_1 = mean(ANESraw[ANESraw$immig_numb == 1, ]$ftsanders)
sanders_2 = mean(ANESraw[ANESraw$immig_numb == 2, ]$ftsanders)
sanders_3 = mean(ANESraw[ANESraw$immig_numb == 3, ]$ftsanders)
sanders_4 = mean(ANESraw[ANESraw$immig_numb == 4, ]$ftsanders)
sanders_5 = mean(ANESraw[ANESraw$immig_numb == 5, ]$ftsanders)
sanders_6 = mean(ANESraw[ANESraw$immig_numb == 6, ]$ftsanders)
sanders_7 = mean(ANESraw[ANESraw$immig_numb == 7, ]$ftsanders)
sanders = c(sanders_1, sanders_2, sanders_3, sanders_4, sanders_5, sanders_6, sanders_7)
summary(sanders)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   30.10   56.29   60.06   56.59   64.40   64.62
```

```
rubio_1 = mean(ANESraw[ANESraw$immig_numb == 1, ]$ftrubio)
rubio_2 = mean(ANESraw[ANESraw$immig_numb == 2, ]$ftrubio)
rubio_3 = mean(ANESraw[ANESraw$immig_numb == 3, ]$ftrubio)
rubio_4 = mean(ANESraw[ANESraw$immig_numb == 4, ]$ftrubio)
rubio_5 = mean(ANESraw[ANESraw$immig_numb == 5, ]$ftrubio)
rubio_6 = mean(ANESraw[ANESraw$immig_numb == 6, ]$ftrubio)
rubio_7 = mean(ANESraw[ANESraw$immig_numb == 7, ]$ftrubio)
rubio = c(rubio_1, rubio_2, rubio_3, rubio_4, rubio_5, rubio_6, rubio_7)
summary(rubio)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   31.54   41.01   47.80   53.30   55.86  100.04
```

```
subsetData = data.frame(trump,hillary,sanders,rubio)
plot1 = ggplot(subsetData, aes(y = trump, x = seq(1, 7))) + geom_point() + xlab("Immigration Num
ber") + ylab("Feeling Thermometer Mean") + ggtitle("Donald Trump")
plot2 = ggplot(subsetData, aes(y = hillary, x = seq(1, 7))) + geom_point() + xlab("Immigration N
umber") + ylab("Feeling Thermometer Mean") + ggtitle("Hillary Clinton")
plot3 = ggplot(subsetData, aes(y = sanders, x = seq(1, 7))) + geom_point() + xlab("Immigration N
umber") + ylab("Feeling Thermometer Mean") + ggtitle("Bernie Sanders")
plot4 = ggplot(subsetData, aes(y = rubio, x = seq(1, 7))) + geom_point() + xlab("Immigration Num
ber") + ylab("Feeling Thermometer Mean") + ggtitle("Marco Rubio")
grid.arrange(plot1, plot2, plot3, plot4, nrow = 2, ncol = 2)
```




From the graphs plotted above, we can say that 1. There are a lot of supporters of Donald Trump who want the number of immigrants to be reduced by a lot. 2. There are a lot of supporters of Hillary Clinton who want the number of immigrants to be increased by a considerable amount. 3.There are a few supporters of Bernie Sanders who believe that legal immigration should be increased. 4. The supporters of Marco Rubio believe that the rate of legal immigration has to be reduced by a slight amount.