

Homework 08 - STAT440

Joseph Sepich (jps6444)

11/01/2020

```
set.seed(42)
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.5.3
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

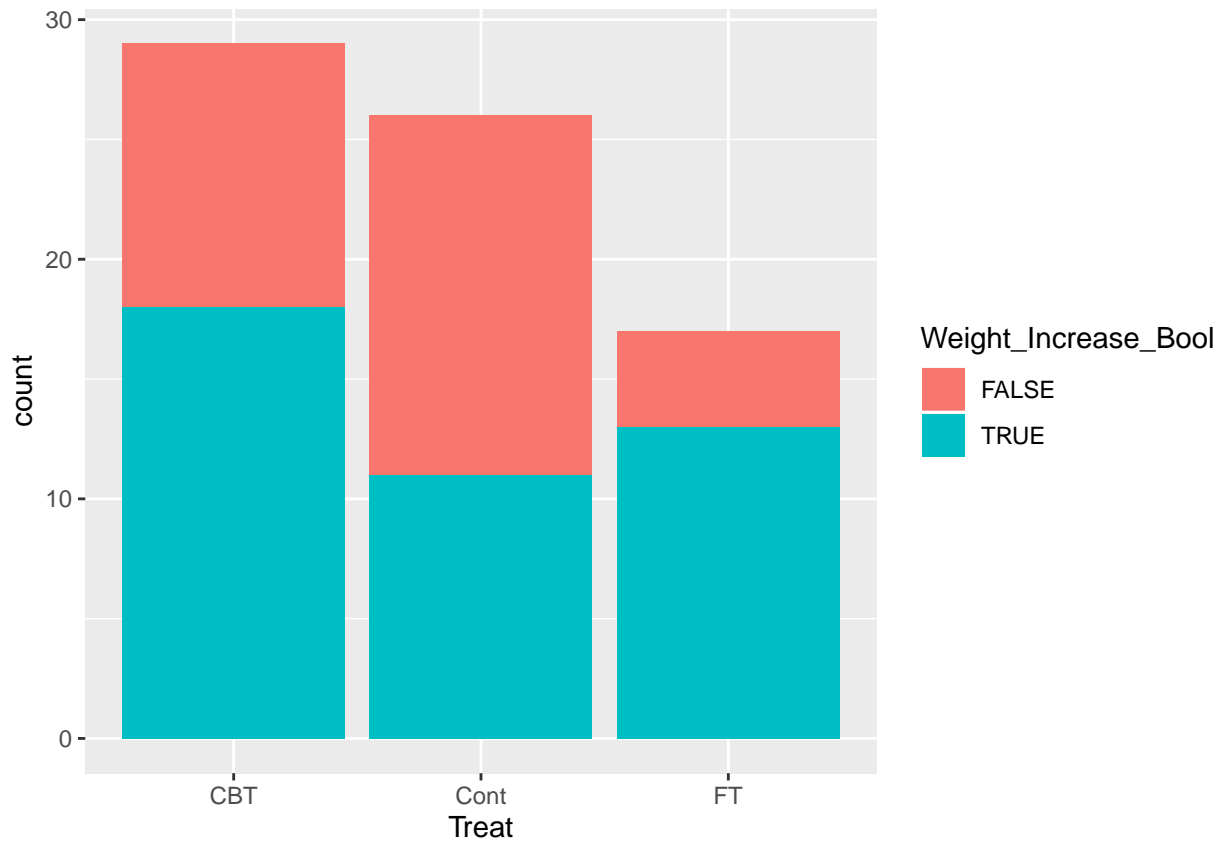
```
##      select
```

```
df <- anorexia
```

Problem 1

```
df$Weight_Increase <- as.numeric(df$Postwt > df$Prewt)
df$Weight_Increase_Bool <- df$Postwt > df$Prewt
```

```
df %>%  
  ggplot(aes(x=Treat)) +  
  geom_bar(aes(fill=Weight_Increase_Bool))
```



Problem 2

Simplifying our data into a simply binary variable removes the weight/value of the data that comes with analyzing the difference in weight from pre to post treatment. This could be both good and bad. Sometimes we like to use this information to help quantify how significant of a change occurs; however since we are going to be looking between groups, and mostly care if one group had a significant increase versus another, we care more about comparing the amount of change between groups. In this way removing the value of differences of each person is good, because people are all different, so the extremity of change can vary a lot, so analyzing with this binary variable will not be greatly affected by outliers.

Problem 3

Problem 4

Problem 5

Problem 6