

## Anatomy of a ggplot

Meredith Dodd <meredith.dodd@woz-u.com>

Mon 5/11/2020 4:12 PM

To: Meredith Dodd <meredith.dodd@woz-u.com>

# Create a histogram with suitable bin widths for the number of absences each student has

```
library("ggplot2")
```

```
ggplot(student.mat, aes(x=absences)) +  
  geom_histogram(bins=50) +  
  ggtitle("Number of Student Absences") +  
  xlab("Absences") + ylab("Frequency Count")
```

# Create a boxplot of the number of absences each student has

```
ggplot(student.mat, aes(y=absences)) +  
  geom_boxplot(outlier.color="red") +  
  ggtitle("Number of Student Absences") +  
  xlab("Absences") + ylab("Frequency Count")
```

# Are there outliers? High or low?

# Yes there are outliers- they are all high

# Create a normal probability plot

```
ggplot(student.mat, aes(sample=absences)) +  
  geom_qq() +  
  ggtitle("Number of Student Absences") +  
  xlab("Absences") + ylab("Frequency Count")
```

# Is the data normally distributed?

#No, because a) you can see a good chunk of the line even if you put a fat pencil from the first dot to the last dot

# and b) because if you go back and look at the histogram, it doesn't look like the bell curve

**Meredith Dodd, Ph.D. | Data Science Program Chair and Instructor**

[meredith.dodd@woz-u.com](mailto:meredith.dodd@woz-u.com)

o: 480-291-8068

powered by



<https://woz-u.com>