

Basic Plots

Eric Hare, Andee Kaplan, Sam Tyner

May 17, 2016

Let's Run the Setup File...

You should see a plot appear if setup is successful.

ggplot2 In a Nutshell

- Package for statistical graphics
- Developed by Hadley Wickham (An ISU Alumni)
- Designed to adhere to good graphical practices
- Supports a wide variety plot types
- Constructs plots using the concept of layers
- <http://had.co.nz/ggplot2/> or Hadley's book *ggplot2: Elegant Graphics for Data Analysis* for reference material

qplot Function

The *qplot()* function is the basic workhorse of ggplot2

- Produces all plot types available with ggplot2
- Allows for plotting options within the function statement
- Creates an object that can be saved
- Plot layers can be added to modify plot complexity

qplot Structure

The *qplot()* function has a basic syntax:

qplot(variables, plot type, dataset, options)

- variables: list of variables used for the plot
- plot type: specified with a *geom* = statement
- dataset: specified with a *data* = statement
- options: there are so, so many options!

Diamonds Data

We will explore the diamonds data set (preloaded along with ggplot2) using qplot for basic plotting.

The data set was scraped from a diamond exchange company data base by Hadley. It contains the prices and attributes of over 50,000 diamonds

Examining the Diamonds Data

What does the data look like?

Lets look at the top few rows of the diamond data frame to find out!

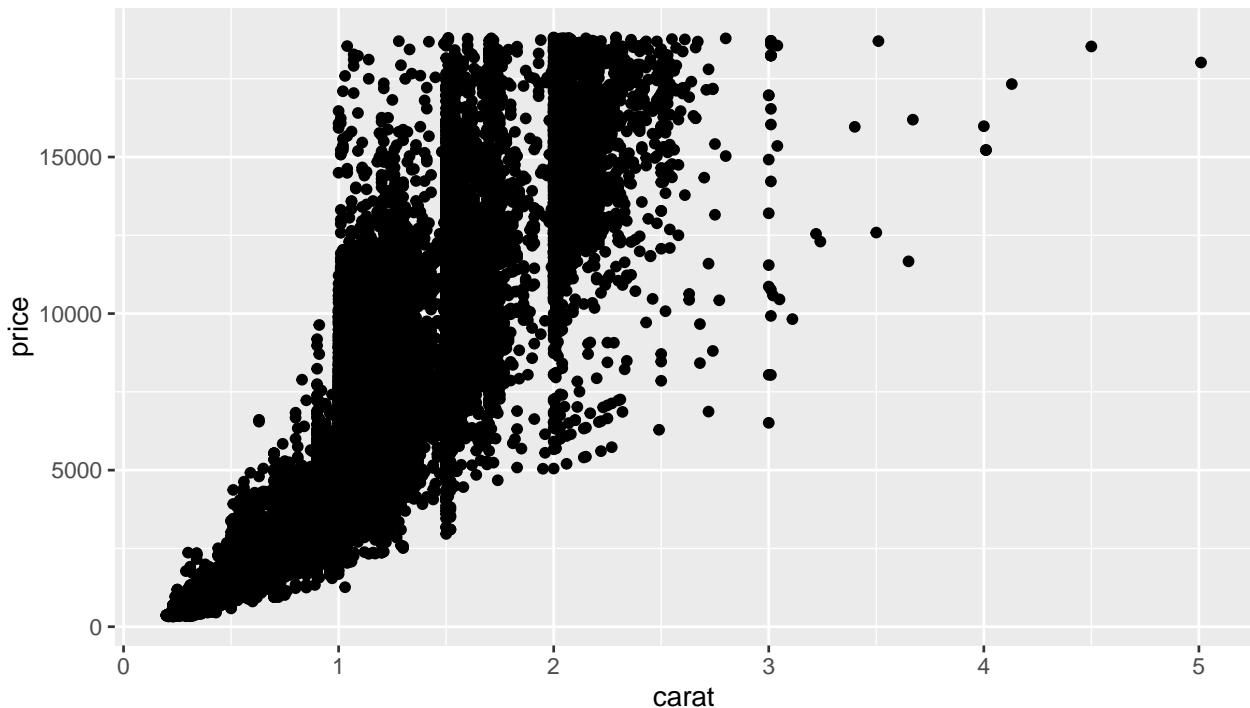
```
head(diamonds)
```

```
##   carat      cut color clarity depth table price     x     y     z
## 1  0.23    Ideal    E    SI2  61.5     55   326 3.95 3.98 2.43
## 2  0.21  Premium    E    SI1  59.8     61   326 3.89 3.84 2.31
## 3  0.23      Good    E    VS1  56.9     65   327 4.05 4.07 2.31
## 4  0.29  Premium    I    VS2  62.4     58   334 4.20 4.23 2.63
## 5  0.31      Good    J    SI2  63.3     58   335 4.34 4.35 2.75
## 6  0.24  Very Good    J   VVS2  62.8     57   336 3.94 3.96 2.48
```

Basic Scatterplot

Basic scatter plot of diamond price vs carat weight

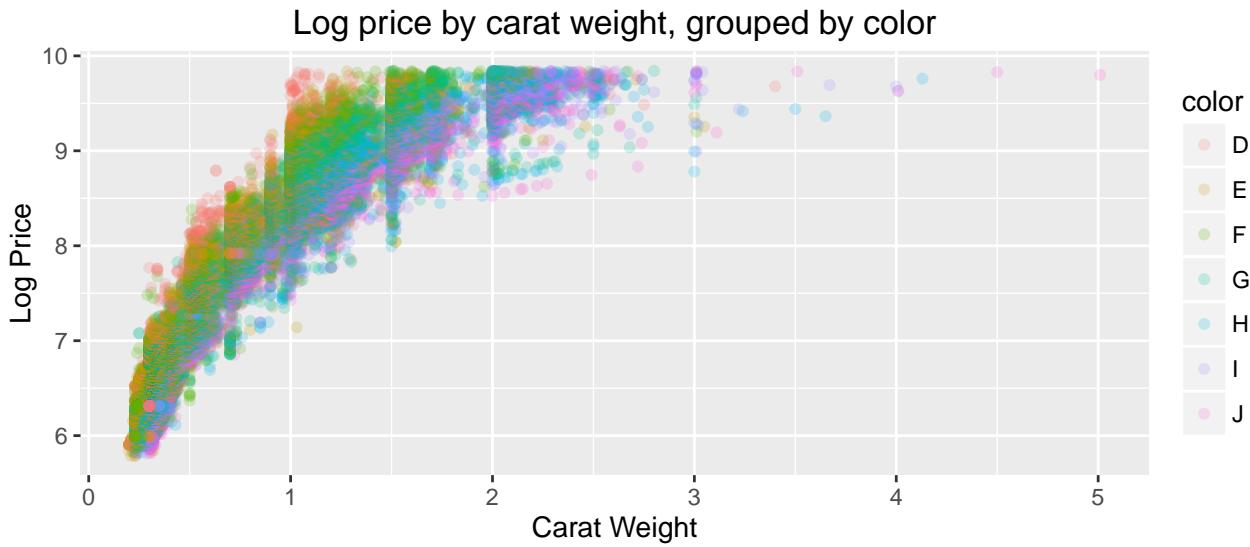
```
qplot(carat, price, geom = "point", data = diamonds)
```



Another Scatterplot

Scatter plot of diamond price vs carat weight showing versility of options in qplot

```
qplot(carat, log(price), geom = "point", data = diamonds,
      alpha = I(0.2), colour = color,
      main = "Log price by carat weight, grouped by color") +
      xlab("Carat Weight") + ylab("Log Price")
```



Your Turn

All of the your turns for this section will use the tips data set:

```
tips <- read.csv("http://heike.github.io/rwrks/02-r-graphics/data/tips.csv")
```

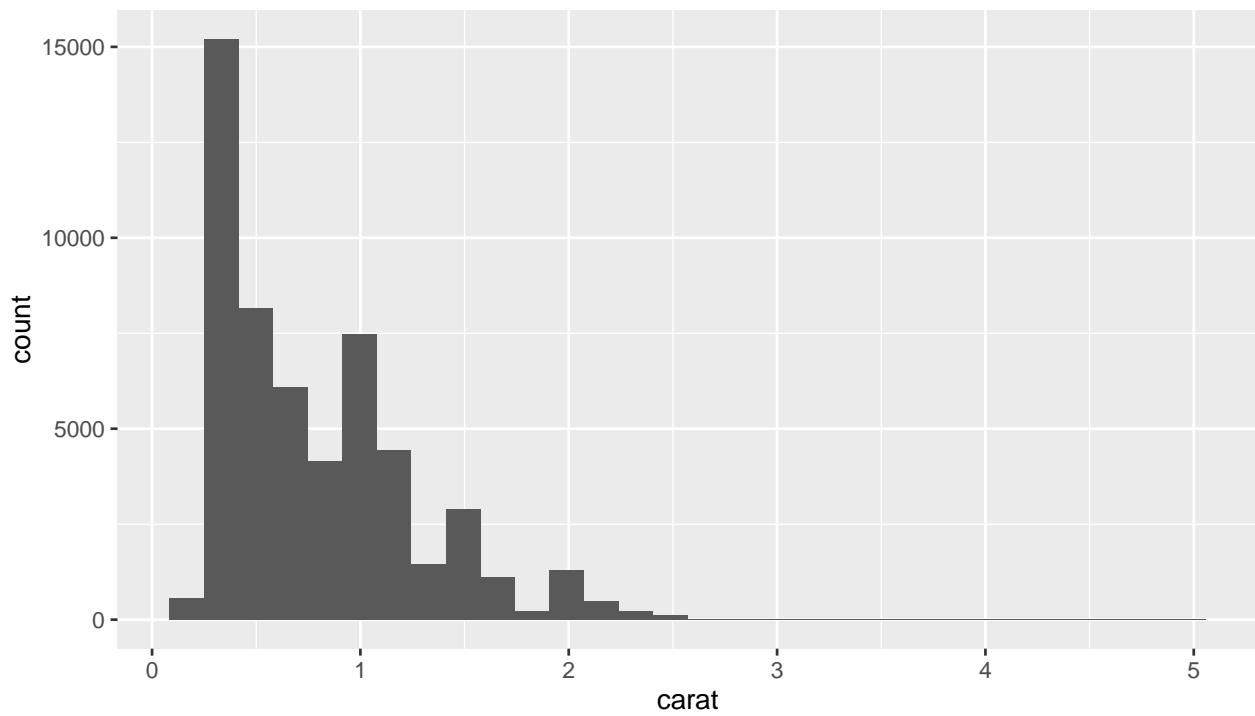
1. Use qplot to build a scatterplot of variables tips and total bill
2. Use options within qplot to color points by smokers
3. Clean up axis labels and add main plot title

Basic Histogram

Basic histogram of carat weight

```
qplot(carat, geom = "histogram", data = diamonds)
```

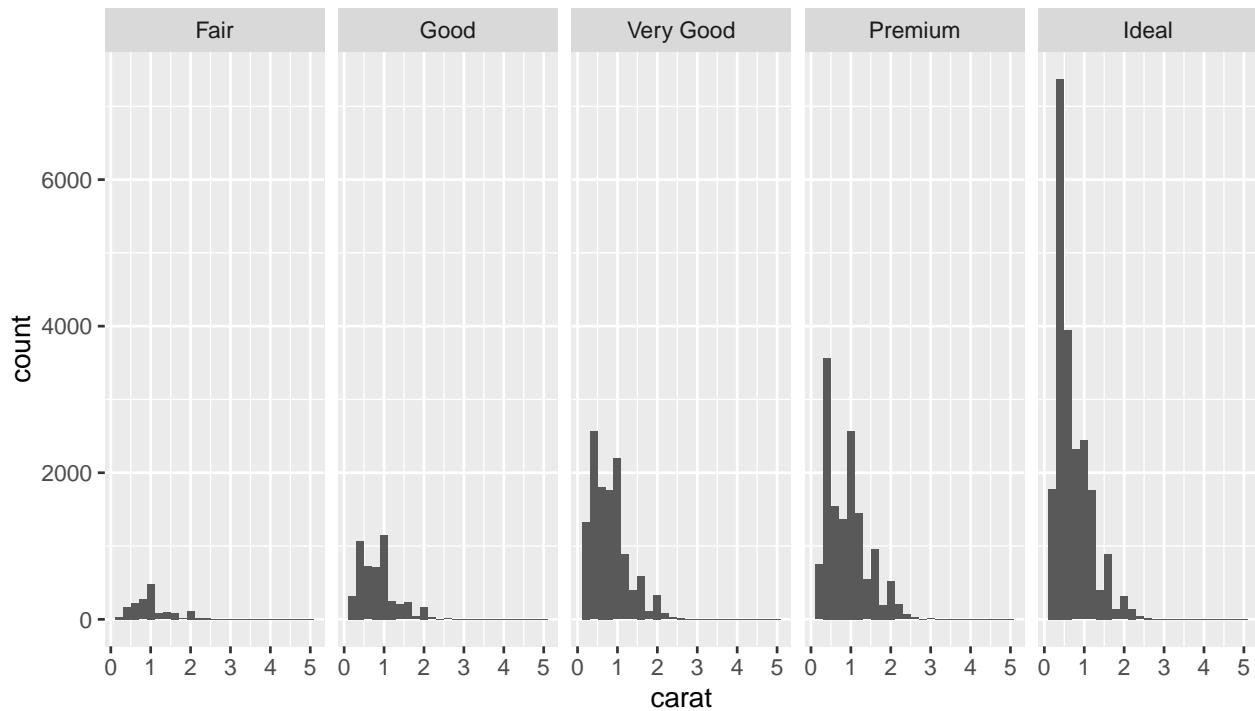
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Another Histogram

Carat weight histograms faceted by cut

```
qplot(carat, geom = "histogram", data = diamonds, binwidth = 0.2, facets = .~cut)
```



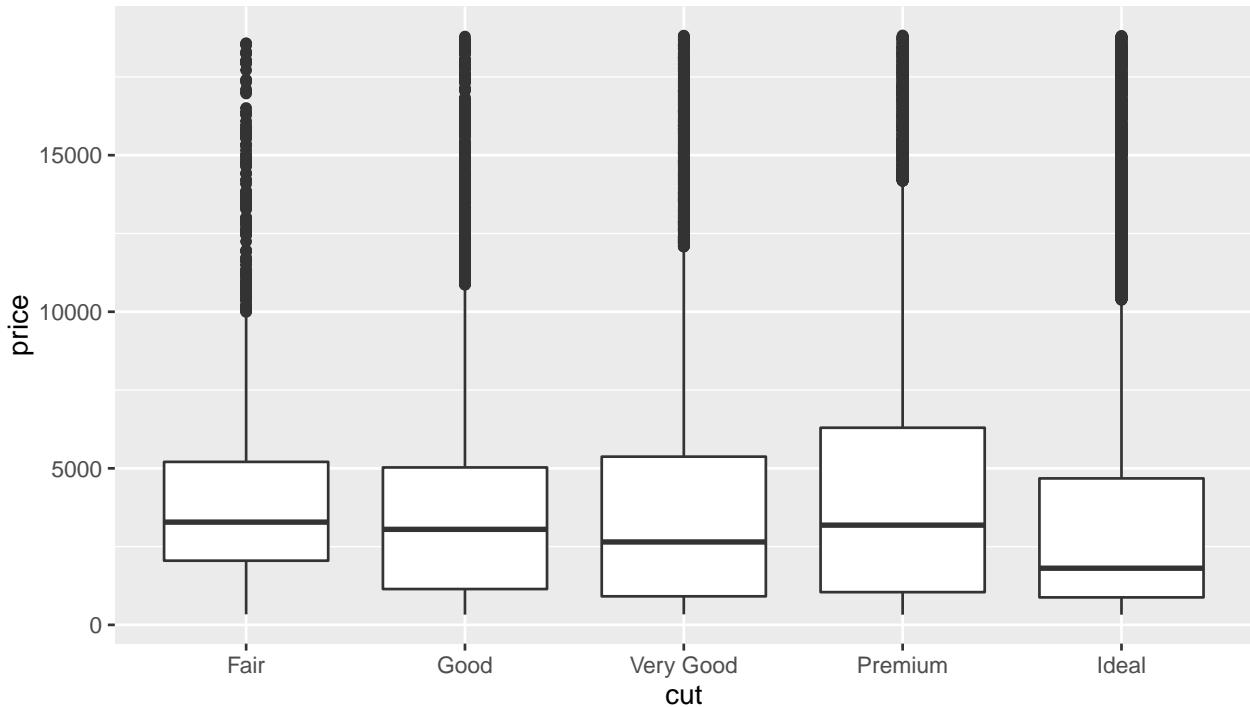
Your Turn

1. Create a new variable in tips data frame rate = tip / total bill
2. Use qplot to create a histogram of rate
3. Change the bin width on that histogram to 0.05
4. Facet this histogram by size of the group

Basic Boxplot

Side by side boxplot of diamond prices within cut groupings

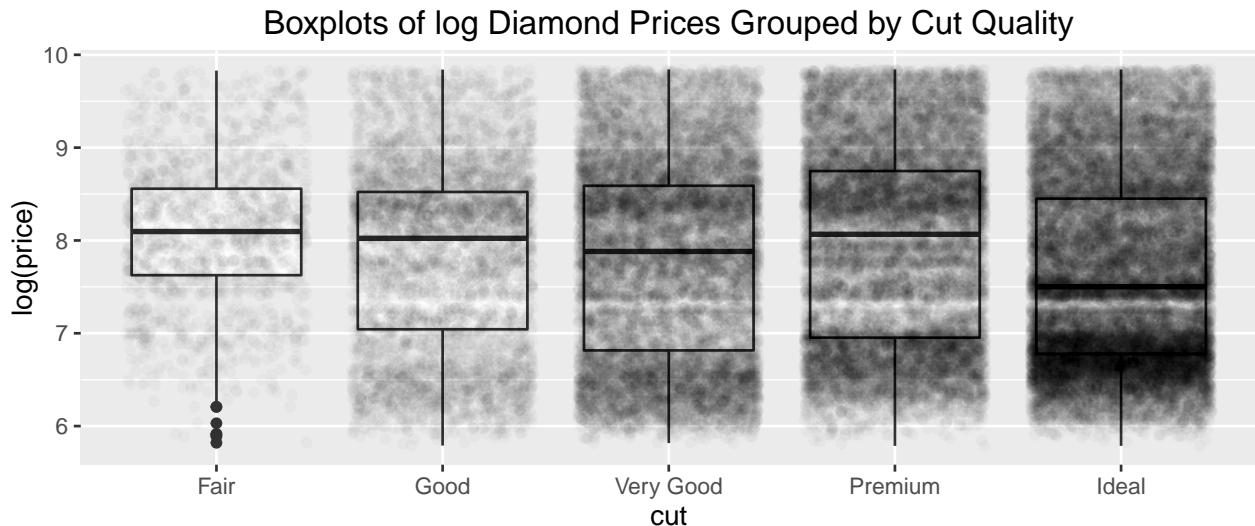
```
qplot(cut, price, geom = "boxplot", data = diamonds)
```



Another Boxplot

Side by side boxplot of log prices within cut groupings with jittered values overlay

```
qplot(cut, log(price), geom = "boxplot", data = diamonds,
      main = "Boxplots of log Diamond Prices Grouped by Cut Quality") +
      geom_jitter(alpha = I(.025))
```



Your Turn

1. Make side by side boxplots of tipping rate for males and females
2. Overlay jittered points for observed values onto this boxplot

Bar Plots

To investigate bar plots we will switch over to the Titanic data set:

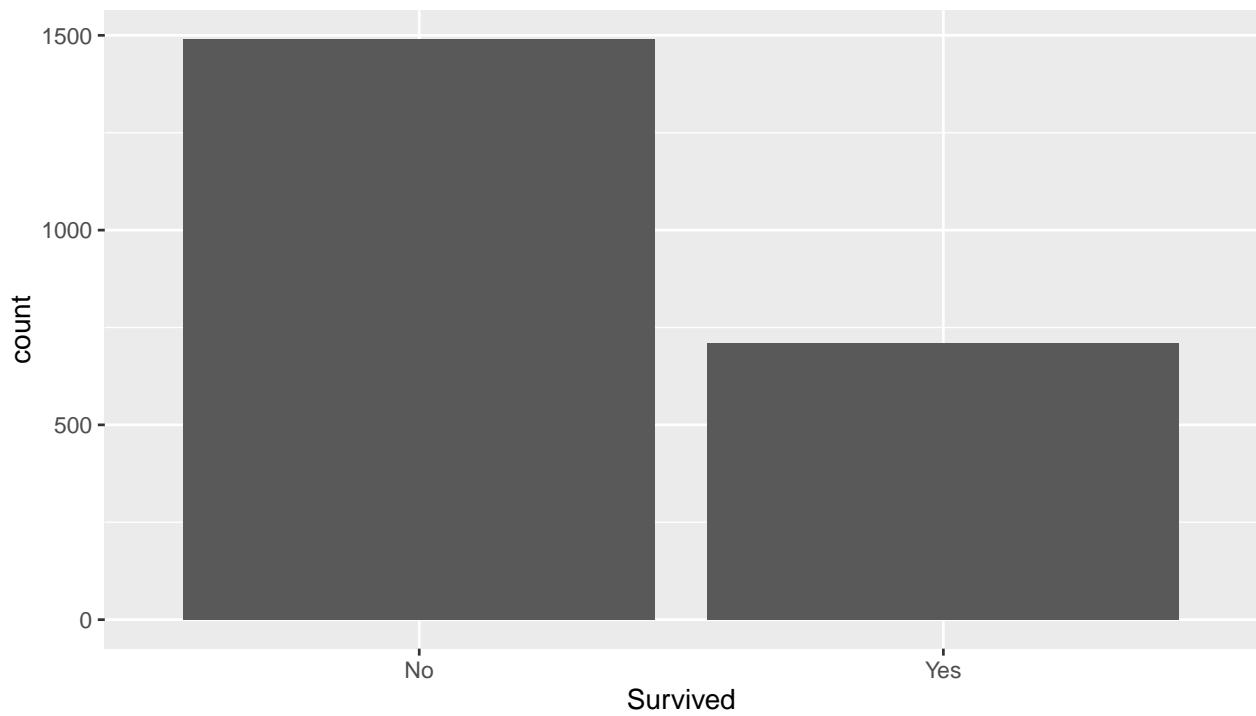
```
titanic <- as.data.frame(Titanic)
```

Data includes passenger characteristics and survival outcomes for those aboard the RMS Titanic's ill fated maiden voyage

Basic Bar Plot

Basic bar plot of survival outcomes

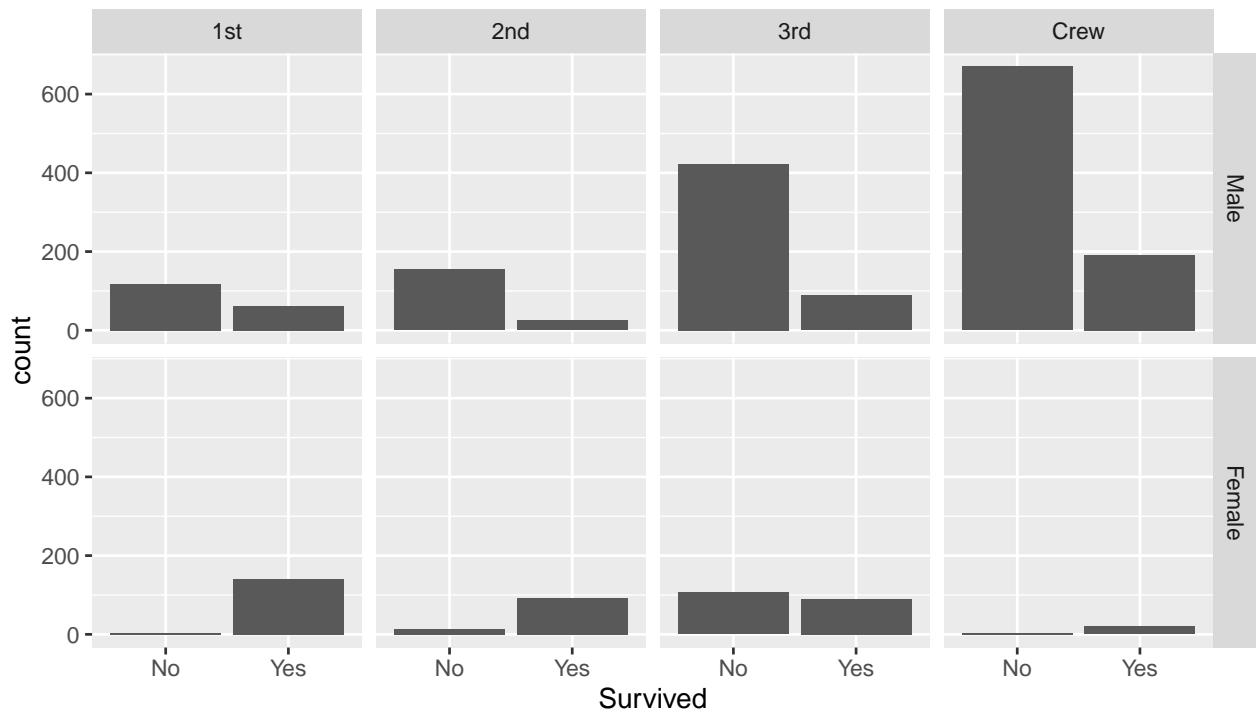
```
qplot(Survived, geom = "bar", data = titanic, weight = Freq)
```



Another Bar Plot

Bar plot faceted by gender and class

```
qplot(Survived, geom = "bar", data = titanic, weight = Freq,
      facets = Sex~Class)
```



Your Turn

1. Use the tips data to make a barplot for counts of smoking and non smoking customers
2. Facet using day of week and time of day to view how smoking status changes for different meal times