Introduction to R

Basic Graphics

ggplot2

- We will use the ggplot2 package for graphics.
- Whenever you first want to plot during an R session you will need to enter the command:

```
library(ggplot2)
```

- If you do not have ggplot installed, you can install the package in the R session by entering the command:

```
install.packages("ggplot2")
```

- A reference manual can be found on Hadley Wickham's website http://had.co.nz/ggplot2/

qplot

- The workhorse function for basic plots in ggplot2 is qplot and has basic usage

```
qplot(x, ..., data, geom)
```

  - x: values of one variable of interest
  - data: data frame
  - geom: the type of plot to construct
- For additional help and examples use the command

```
?qplot
```


Diamonds Data

- ~54,000 round diamonds from http://www.diamondse.info/
- Additional information on carat, color, clarity, and cut
- For more information use the command

```
?diamonds
```

What can we learn from this data?

- Inspect the data
- Figure out what the variables are from http://www.diamondse.info/ and wikipedia
- Think of questions that you could answer with this data

Bar Charts

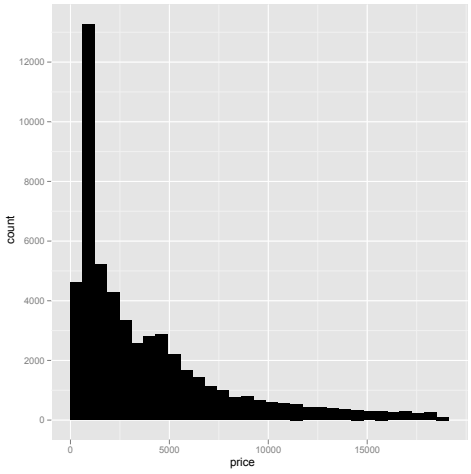
- Explore how one (or more) categorical variables are distributed.
- `qplot(cut, data = diamonds, geom = "bar")`

Histograms

- Explore how one (or more) quantitative variables are distributed.
- `qplot(price, data = diamonds, geom = "histogram", binwidth = 10)`
- `qplot(price, data = diamonds, geom = "histogram", binwidth = 500)`
- `qplot(price, data = diamonds, geom = "histogram", binwidth = 1000)`

Histograms

- Skewed to the right
  - There are lots of "inexpensive" diamonds and a few very, very expensive diamonds



Your turn

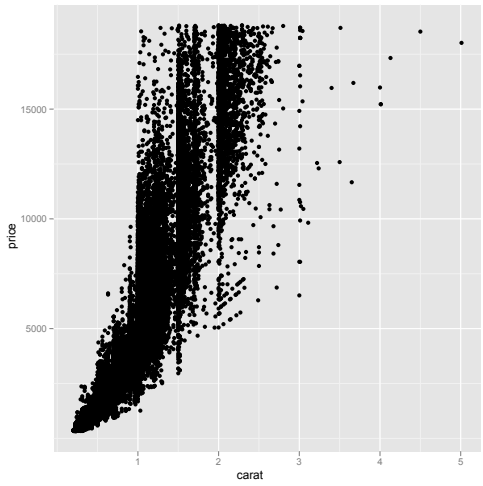
- Create histograms for carat and price per carat
- Experiment with bin width
- Describe what you see

Saving Plots

- There are a few ways to save your plots in R.
  - Using the GUI, choose file → save
  - `png(filename, width, height)`
  - `jpeg(filename, width, height)`
  - `pdf(file, width, height)`
- If the plot has already been rendered, then use
  - `dev.copy(device, filename)`
  - `ggsave(filename, width, height)`

Scatterplots

- Use scatterplots to display the relationship between two quantitative variables.



•Positive nonlinear association of moderate strength
•Variability in the price increases with carat
•There are "gaps" around 1.5 and 2 carats
•There may be outliers above 3 carats

Scatterplots

- ggplot2 will make a scatterplot by default when two variables are entered.
- `qplot(carat, price, data = diamonds)`
- `qplot(log(carat), log(price), data = diamonds)`
- `qplot(carat, price/carat, data = diamonds)`

Adding Smoothers

- `qplot(carat, price, data = diamonds, geom = c("point", "smooth"), method = lm, ylim = c(0, 20000))`
- `qplot(carat, price, data = diamonds, ylim = c(0, 20000)) + geom_smooth (method = lm)`

Adding Aesthetics

- We do not need to ignore the other variables when looking at scatterplots.
- We can map other variables to colour, shape, and size.
- `qplot(carat, price, data = diamonds, colour = color)`
- `qplot(carat, price, data = diamonds, shape = cut)`

Faceting

- Facets allow us to display plots for different subsets.
- facets = row variables ~ column variables (use '.' for none or blank for wrapping)
- `qplot(carat, price, data = diamonds, facets = . ~ color)`
- `qplot(carat, price, data = diamonds, facets = clarity ~ .)`

Boxplots

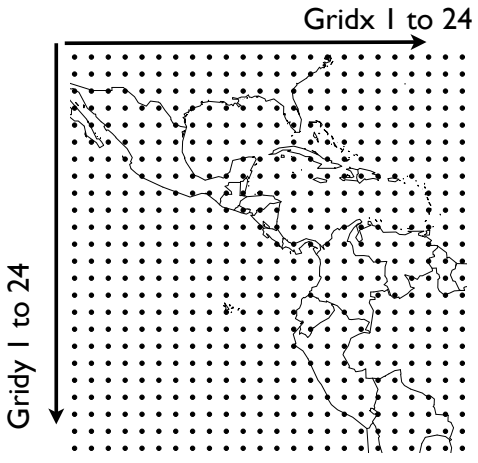
- Useful when we wish to investigate the relationship between a categorical and quantitative variable.
- Displaying side-by-side boxplots allow us to examine conditional distributions
- `qplot(clarity, price/carat, data = diamonds, geom = "boxplot")`

Your turn

- Add smoothers to the facetted scatterplots.
- Explore the relationships between price, carats, color, and clarity.
- What do you learn?

NASA Meteorological Data

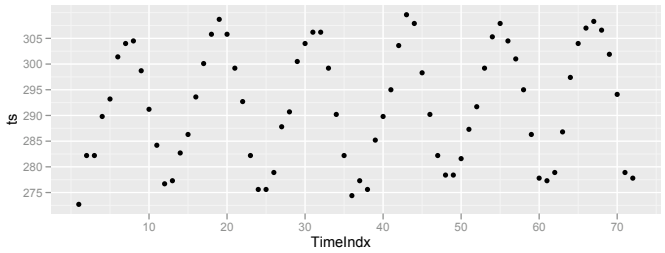
- 24 x 24 grid across Central America
- satellite captured data:
  temperature (ts),
  near surface temperature (tsa)
  pressure (ps)
  ozone (o3)
  cloud coverage:
    low (ca_low)
    medium (ca_med)
    high (ca_high)
- for each location monthly averages for Jan 1995 to Dec 2000



Time Series Plots

for each location multiple measurements

```
qplot(TimeIndx, ts, geom="point",
data=subset(nasa, (Gridx==1) &
(Gridy==1)))
```



connected by a line

```
qplot(TimeIndx, ts, geom="line",
data=subset(nasa, (Gridx==1) &
(Gridy==1)))
```