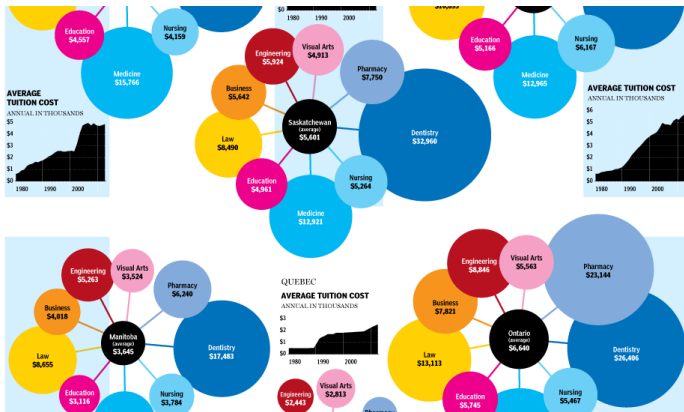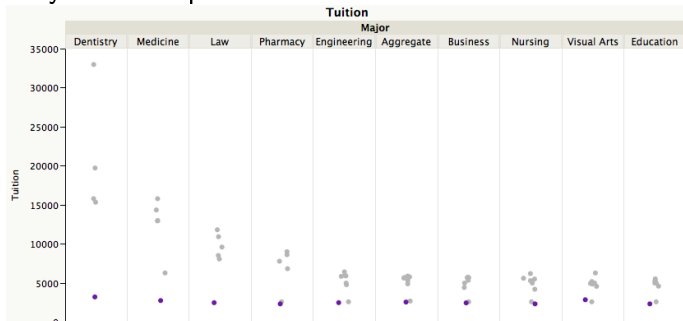# 2 – Advanced Graphics

05 - Perception

Iowa State University

Cost of an Education

Motivation

- Why are some plots easier to read?

- http://junkcharts.typepad.com/junk_charts/2012/05/
  spring-flowers-and-striking-hours.html

Good Graphics

Graphics consist of

- Structure (boxplot, scatterplot, etc.)
- Aesthetics: features such as color, shape, and size that map other characteristics to structural features

Both the structure and aesthetics should help viewers interpret the information.

Outline

- Cognitive aspects of perception and aesthetic choices

- Visual ordering mechanisms and color choices
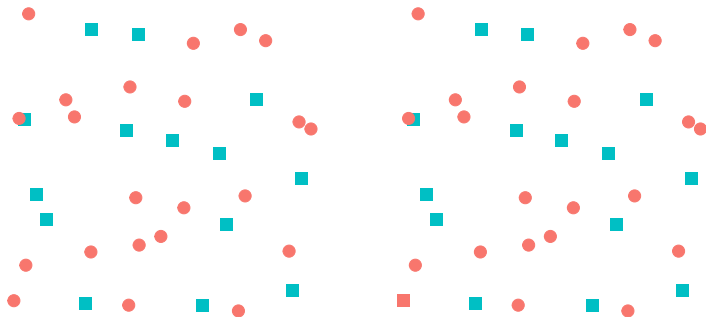
- Faceting graphs to show additional variables

Pre-Attentive Features

- Things that "jump out" in less than 250 ms
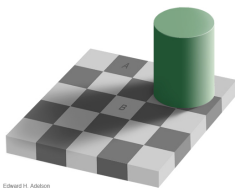
- Color, form, movement, spatial localization
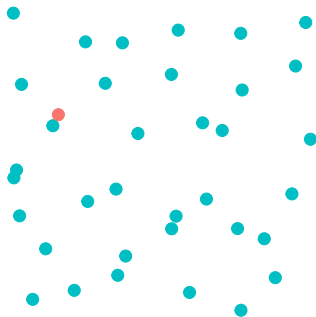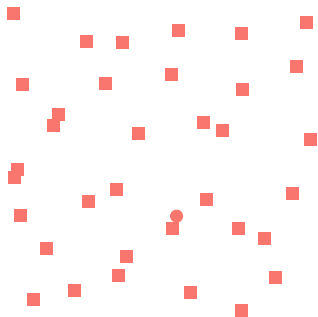
Hierarchy of Features

- Color is stronger than shape

- Combinations of pre-attentive features are usually not pre-attentive due to *interference*

Color

- Hue: shade of color (red, orange, yellow...)
- Intensity: amount of color
- Both color and hue are pre-attentive. Bigger contrast corresponds to faster detection.

Color is context-sensitive: the exact same hue and intensity in one situation may appear to be a different color in a different context. A and B are the same intensity and hue, but appear to be different.

Aesthetics in `ggplot2`

Scales

Scales control the mapping between data and aesthetics. Read more

 scale_alpha
Alpha scale for continuous variable

 scale_brewer
Sequential, diverging and qualitative colour scales from colorbrewer.org

 scale_continuous
Continuous position scale

4/10/197 scale_date
Position scale, date

14/10/197 scale_datetime
10:14am Position scale, date time

scale_discrete
Discrete position scale

 scale_gradient
Smooth gradient between two colours

 scale_gradient2
Smooth gradient between three colours (high, low and midpoints)

 scale_gradientn
Smooth gradient between n colours

scale_grey
Sequential grey colour scale

 scale_hue
Qualitative colour scale with evenly spaced hues

f(x) = x scale_identity
Use values without scaling

scale_linetype
Scale for line patterns

DIY scale_manual
Create your own discrete scale

scale_shape
Scale for shapes, aka glyphs

scale_size
Size scale for continuous variable

Main parameters: alpha, shape, color, size
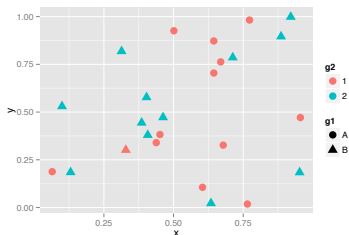
Your Turn

Find ways to improve the following graphic:

```
frame <- data.frame(x=runif(25), y=runif(25),
        g1=rep(c("A","B"), c(12,13)),
        g2=rep(c("1", "2"), c(13,12)))
qplot(x,y, shape=g1, colour=g2, data=frame, size=I(4))
```
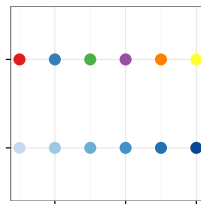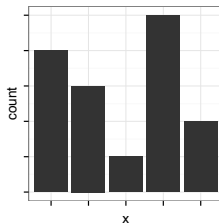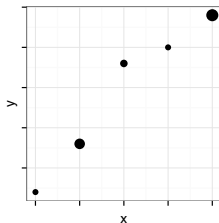
- ▶ Make sure the "oddball"
  stands out while keeping the
  information on the groups
- ▶ Hint: `interaction`
  combines factor variables

Ordering Variables

Which is bigger?

- ▶ Position: higher is bigger (y), items to the right are bigger (x)
- ▶ Size, Area
- ▶ Color: not always ordered. More contrast = bigger.
- ▶ Shape: Unordered.

Using Color

- Qualitative schemes: no more than 7 colors

- Quantitative schemes:
  - use color gradient with only one hue for positive values

  - use color gradient with two hues for positive and negative values. Gradient should go through a light, neutral color (white)
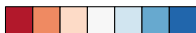
- Small objects or thin lines need more contrast than larger areas

RColorBrewer

R package based on Cynthia Brewer's color schemes
(ColorBrewer2.org)


```
install.packages("RColorBrewer")
library(RColorBrewer)
help(package=RColorBrewer)
display.brewer.all()
```

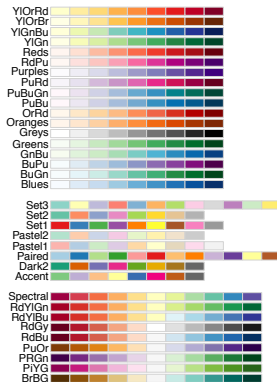# Color in ggplot2

- factor variable:
  `scale_colour_discrete`
  `scale_colour_brewer(palette=...)`
- continuous variable:
  `scale_colour_gradient` (define low, high values)
  `scale_colour_gradient2` (define low, mid, and high values)
- equivalents for fill: `scale_fill_...`

Your Turn

- In the diamonds data, clarity and cut are ordinal, while price and carat are continuous

- Find a graphic that gives an overview of these four variables while respecting their types
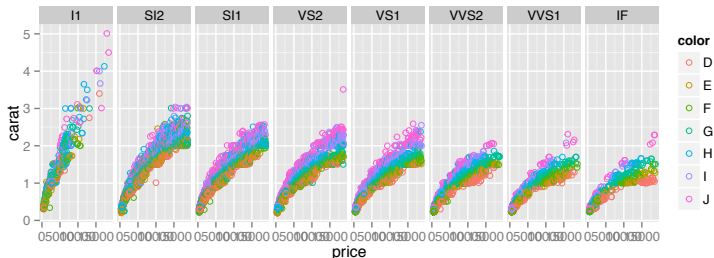
- Hint: Start with

```
data(diamonds)
qplot(carat, price, shape=cut, colour=clarity,
      data=diamonds)
```
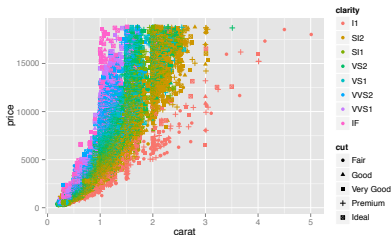
Facetting

- A way to extract subsets of data and place them side-by-side in graphics
- Syntax: `facets = row ~ col` Use `.` if there is no variable for either row or column (i.e. `facets = . ~ col`)

```
qplot(price, carat, data=diamonds, color=color,
    facets = . ~ clarity)
```

Your Turn

- The `movies` dataset contains information from IMDB.com including ratings, genre, length in minutes, and year of release.

- Explore the differences in length, rating, etc. in movie genres over time

- Hint: use facetting!