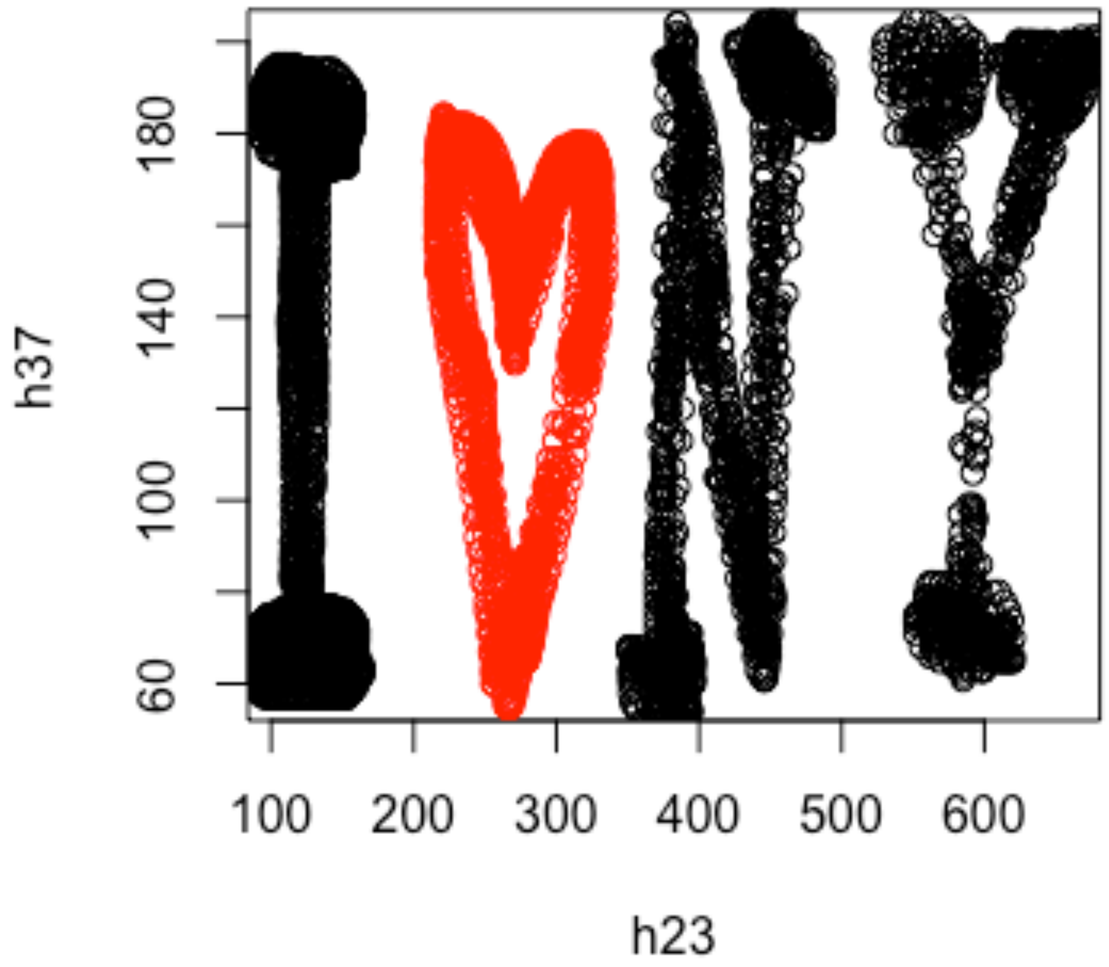
Agenda

- Generating datasets from sketches
- Final Project
- Perception Studies
- Base Graphics

data1.csv



Tools for creating datasets from sketches

- Simple Data Generation

  http://koaning.io/simple-data-generation.html
- Skample!

  https://colcarroll.github.io/skample/
- Drawing Distributions

  http://www.rand-on.com/projects/2017_distribution/distribution.html

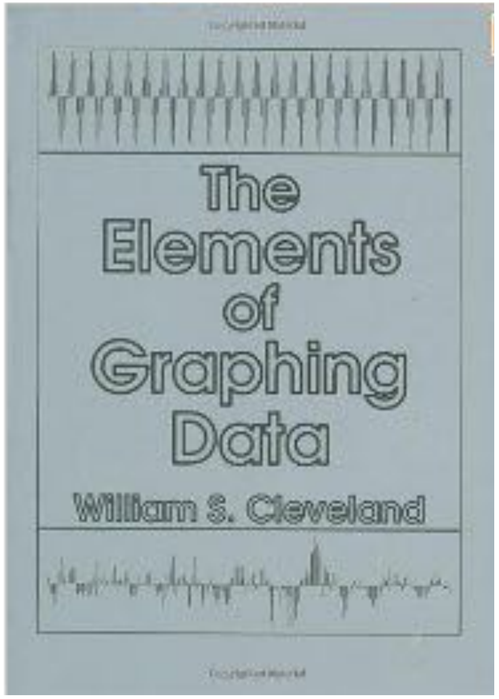
Getting Started on Final Project

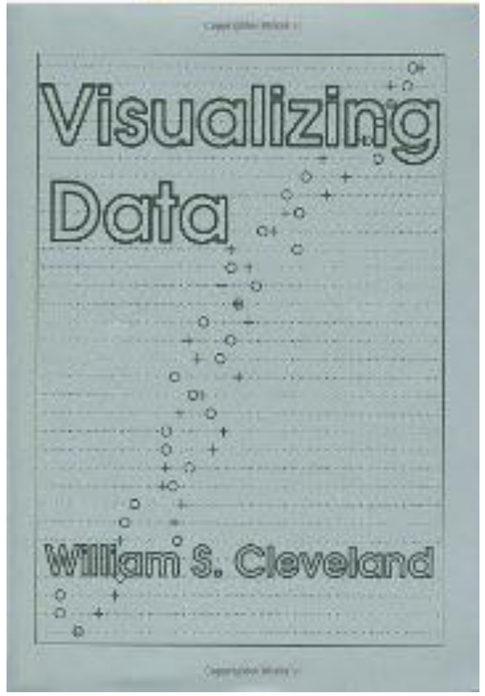
- 2 people per group
- Step #1: choose a topic that interests you
- Step #2: find a dataset that hasn't been overanalyzed
- Step #3: post your idea on the Discussion thread in CourseWorks to find team members

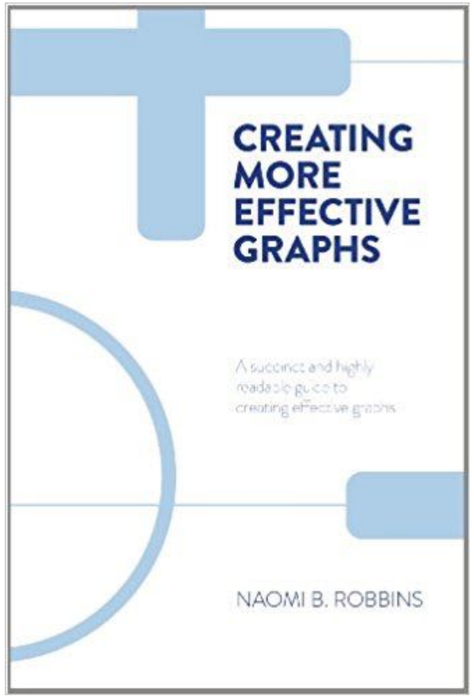
Graphical Perception

articles by William Cleveland and Robert McGill

books by William Cleveland



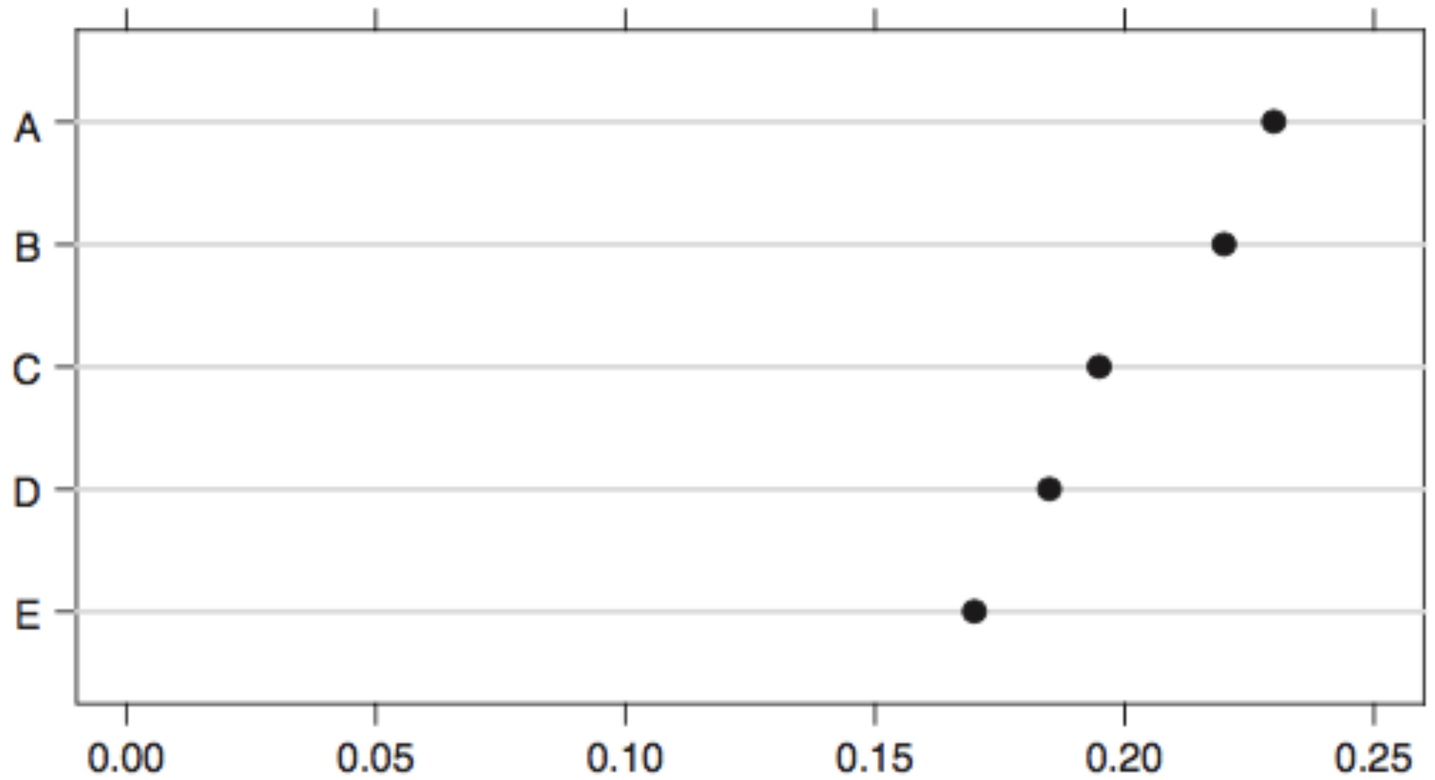




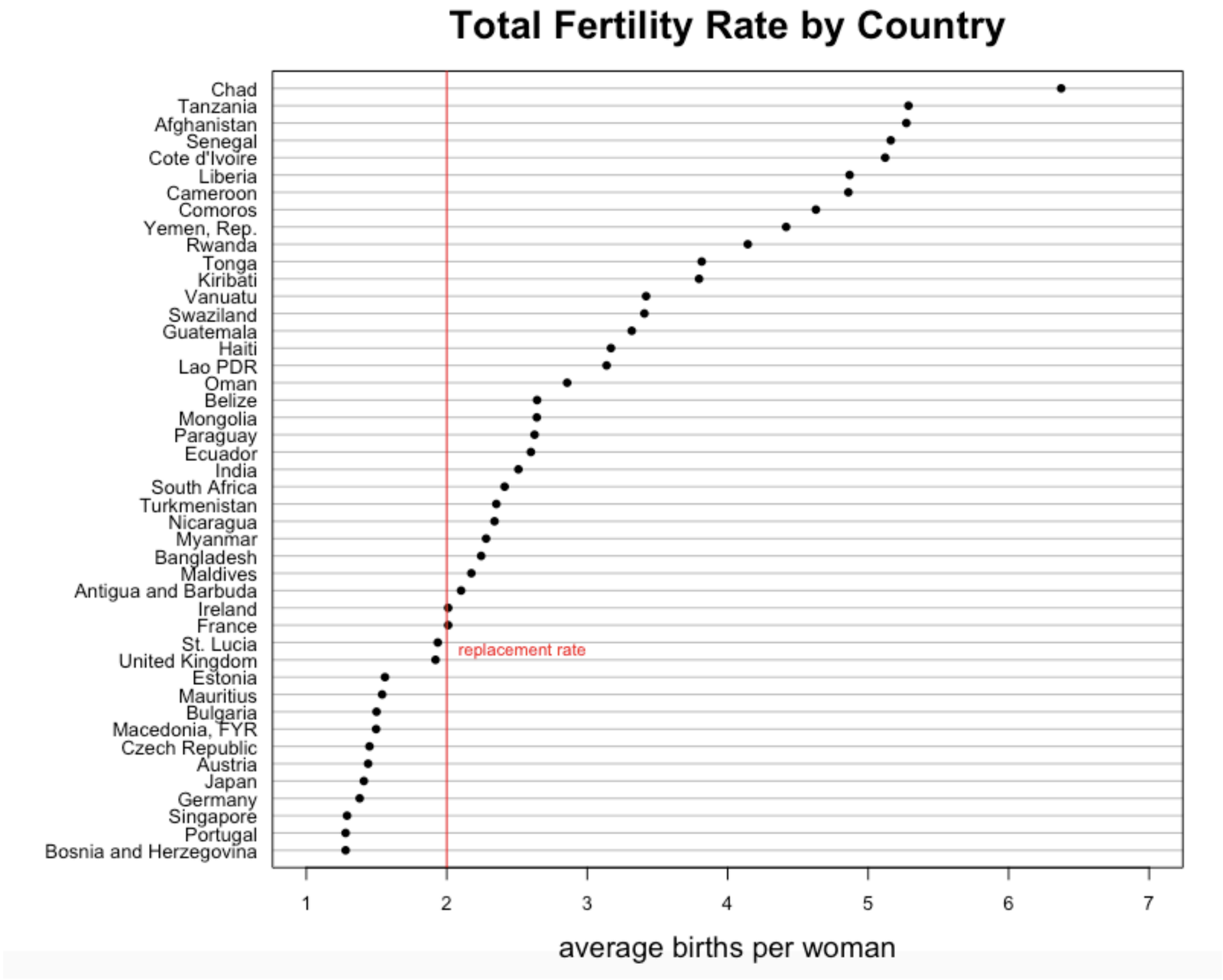
Ordered Elementary Tasks

1. Position along a common scale
2. Position along identical, nonaligned scales
3. Length
4. Angle / Slope
5. Area
6. Volume
7. Color hue / Color saturation / Density

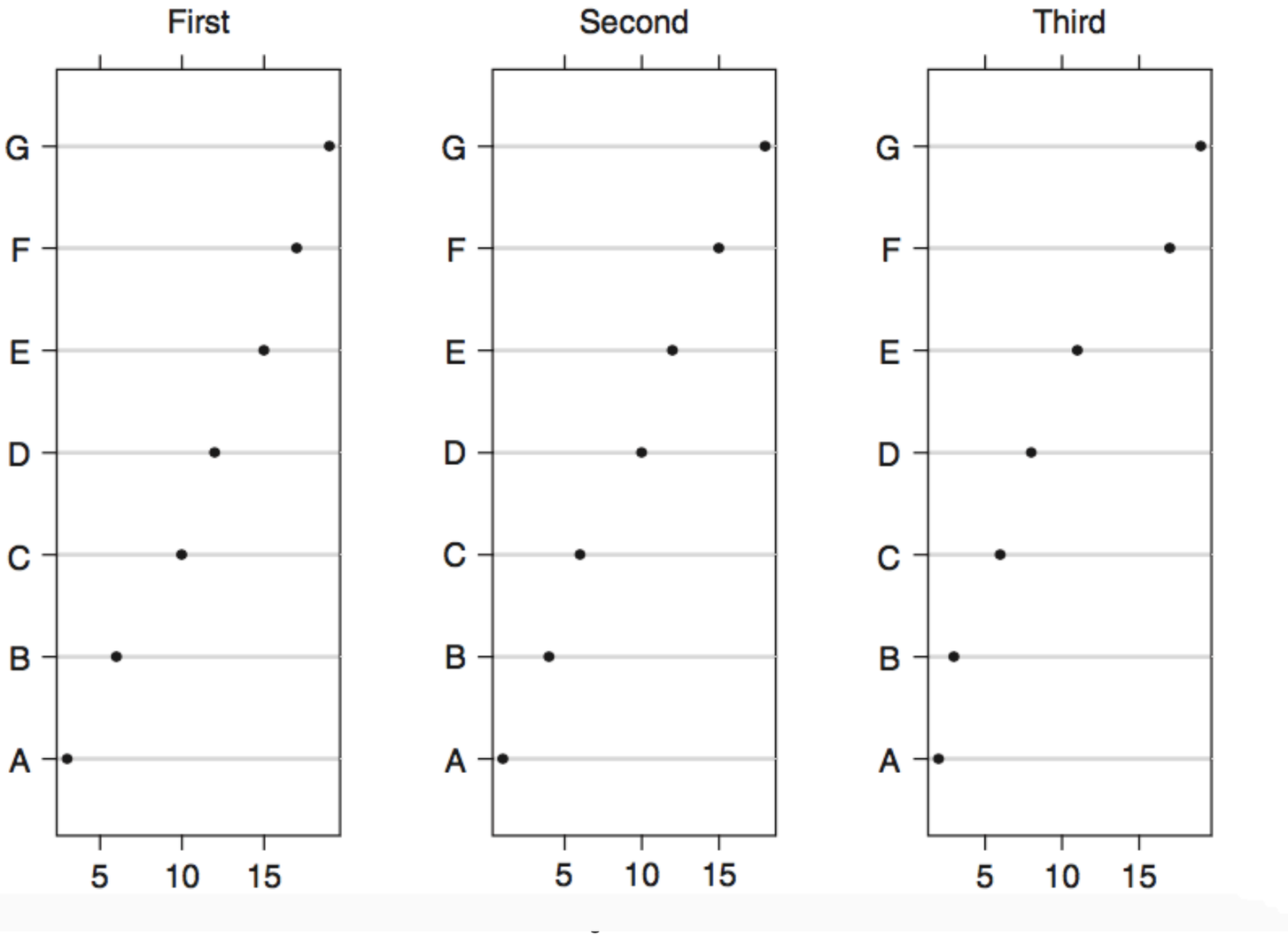
Position along a common scale



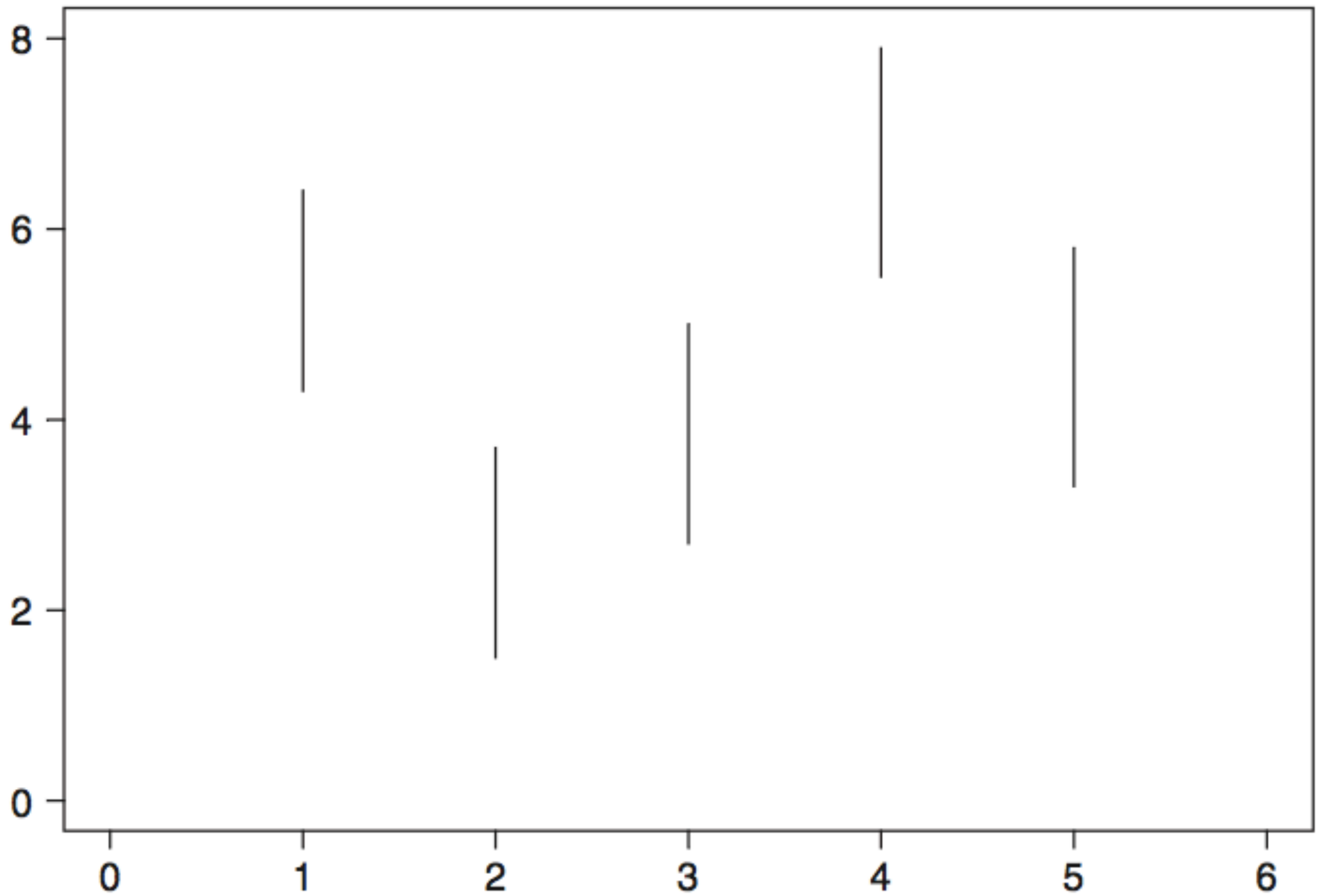
Cleveland dot plot



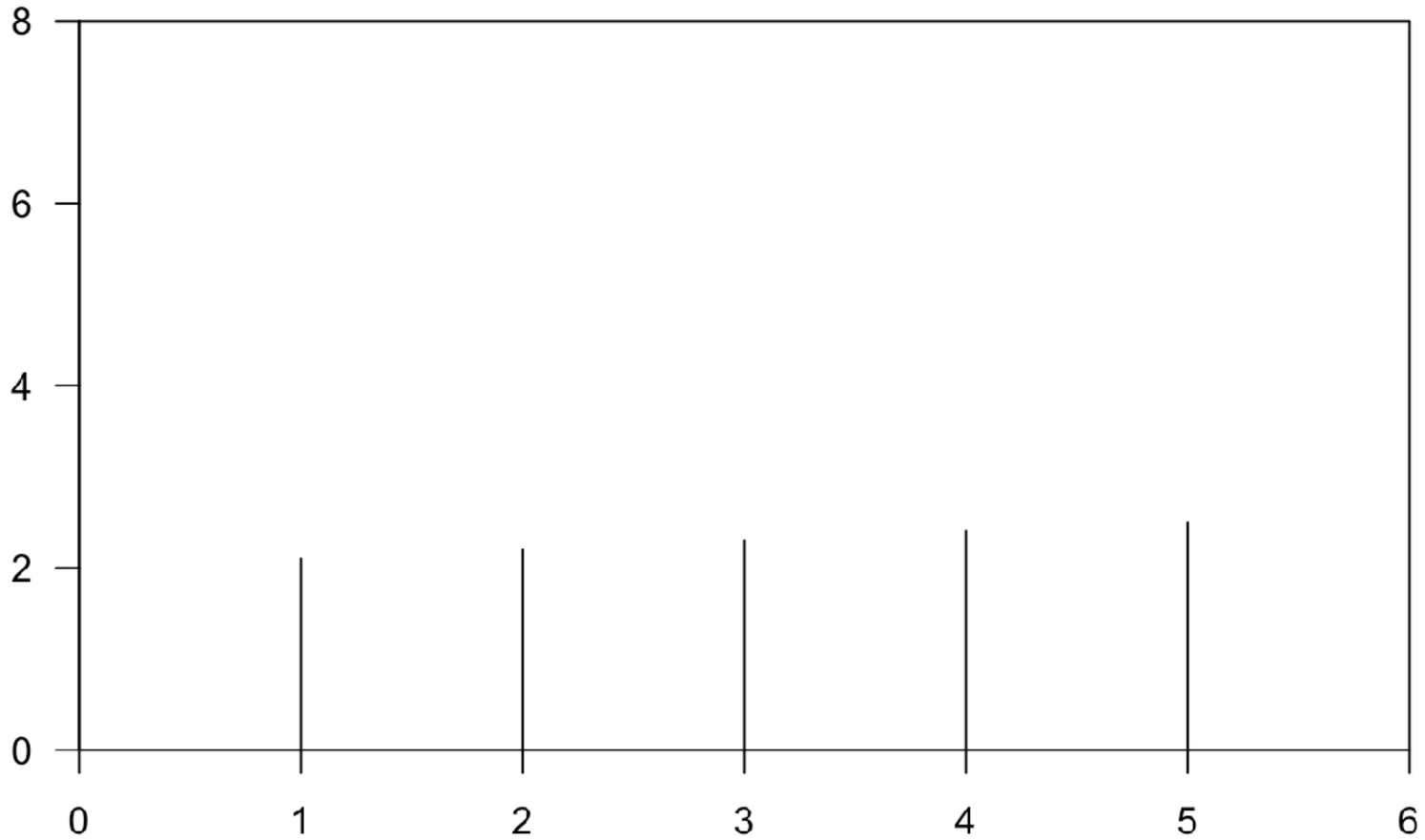
Position along identical, nonaligned scales



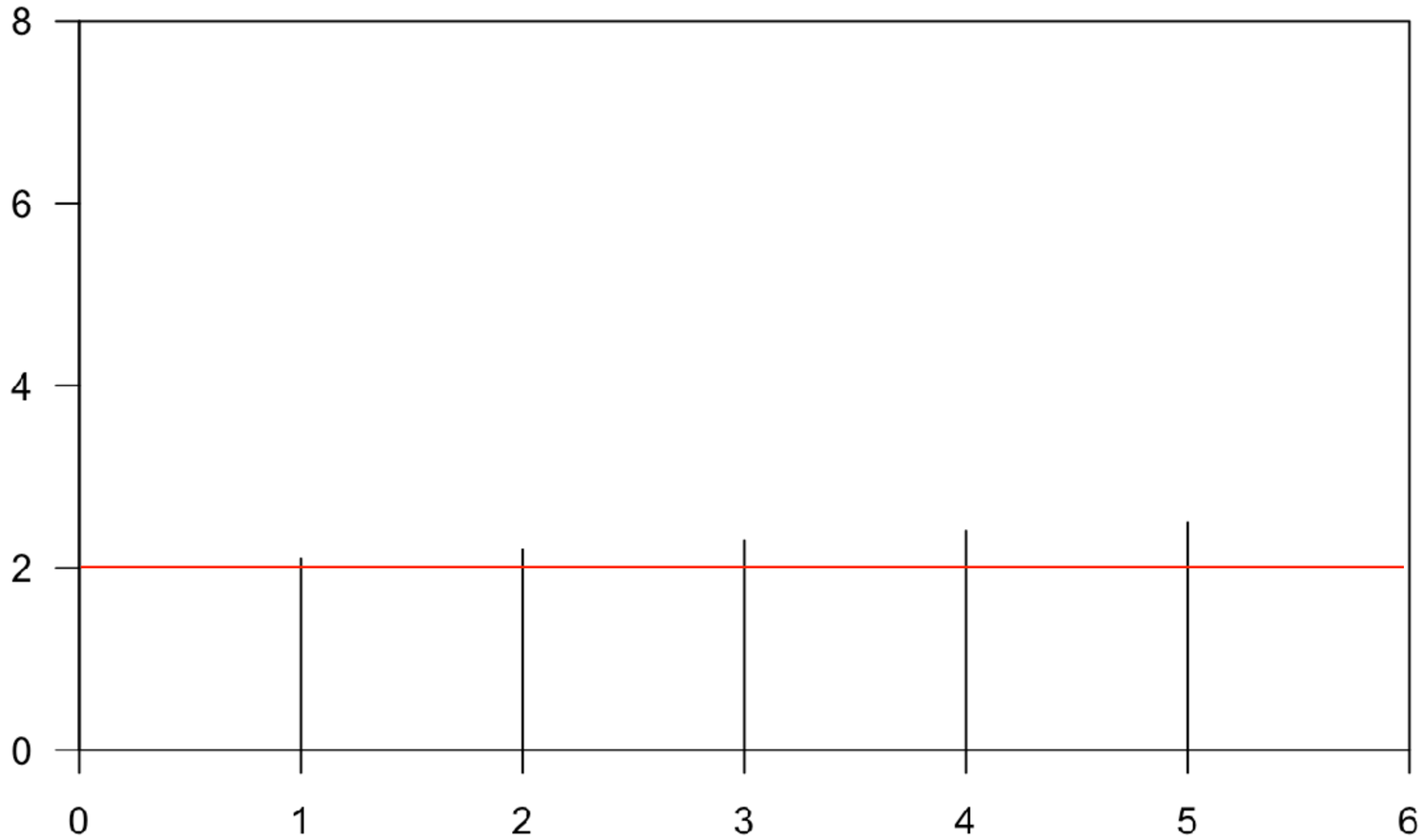
Length



Length



Length



Weber's Law

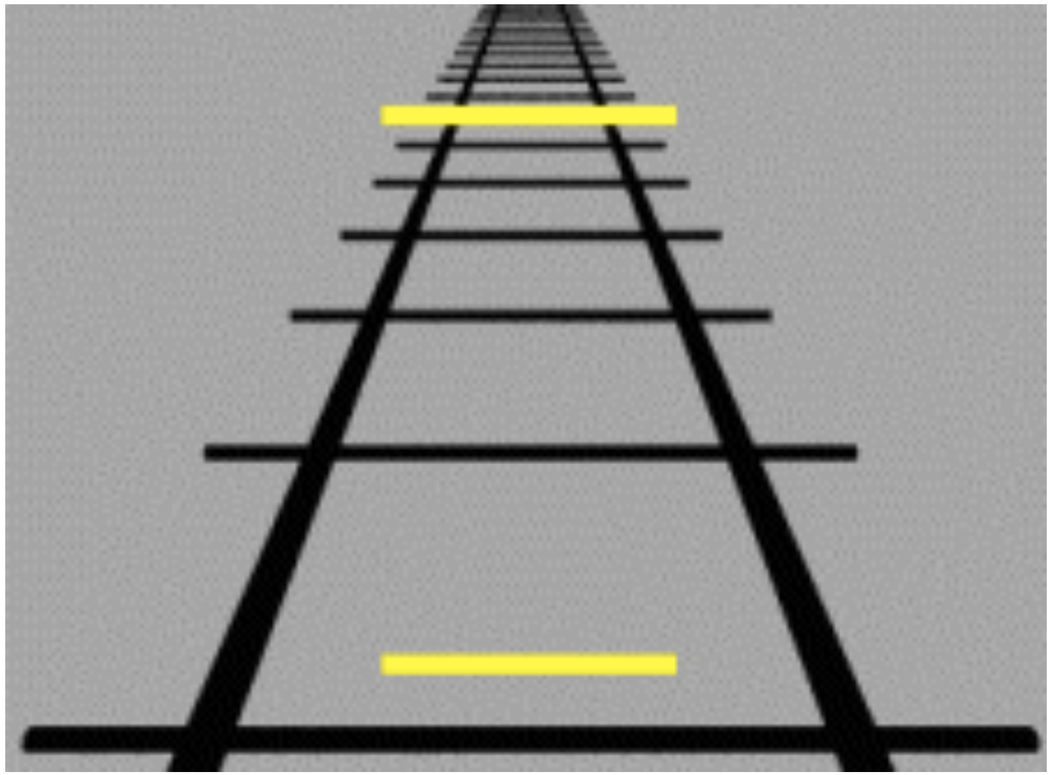
• Consider two lines with lengths x and x + w.

• If w is very small, there is only a very small chance that we will notice that the lines have different lengths.

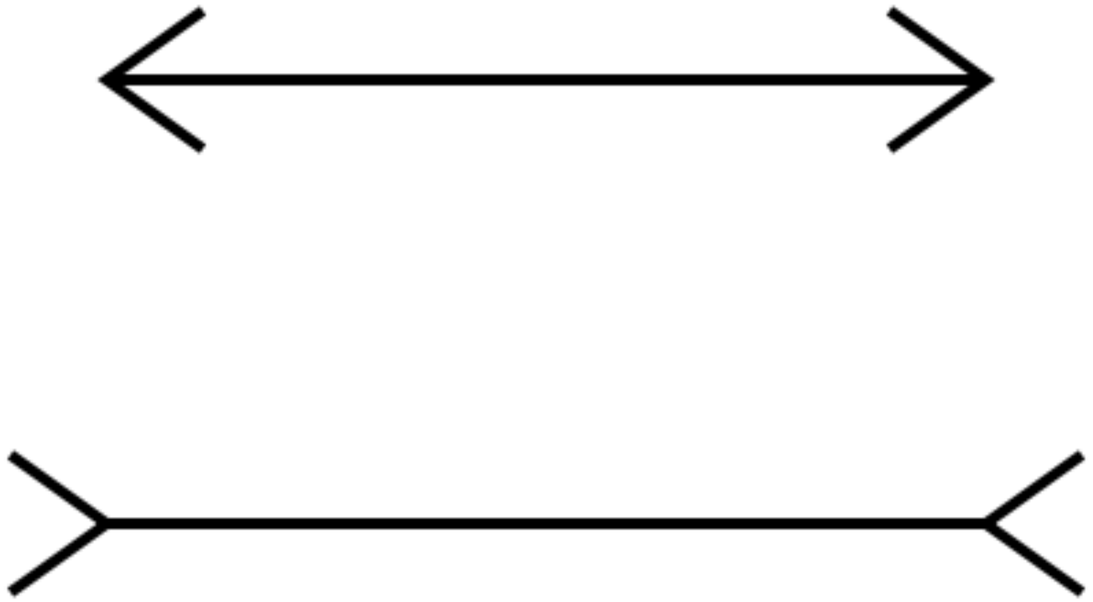
• As w gets larger, the chance of detecting a difference increases.

• Weber’s law says that the chance of detecting a difference depends on the value of w/x.

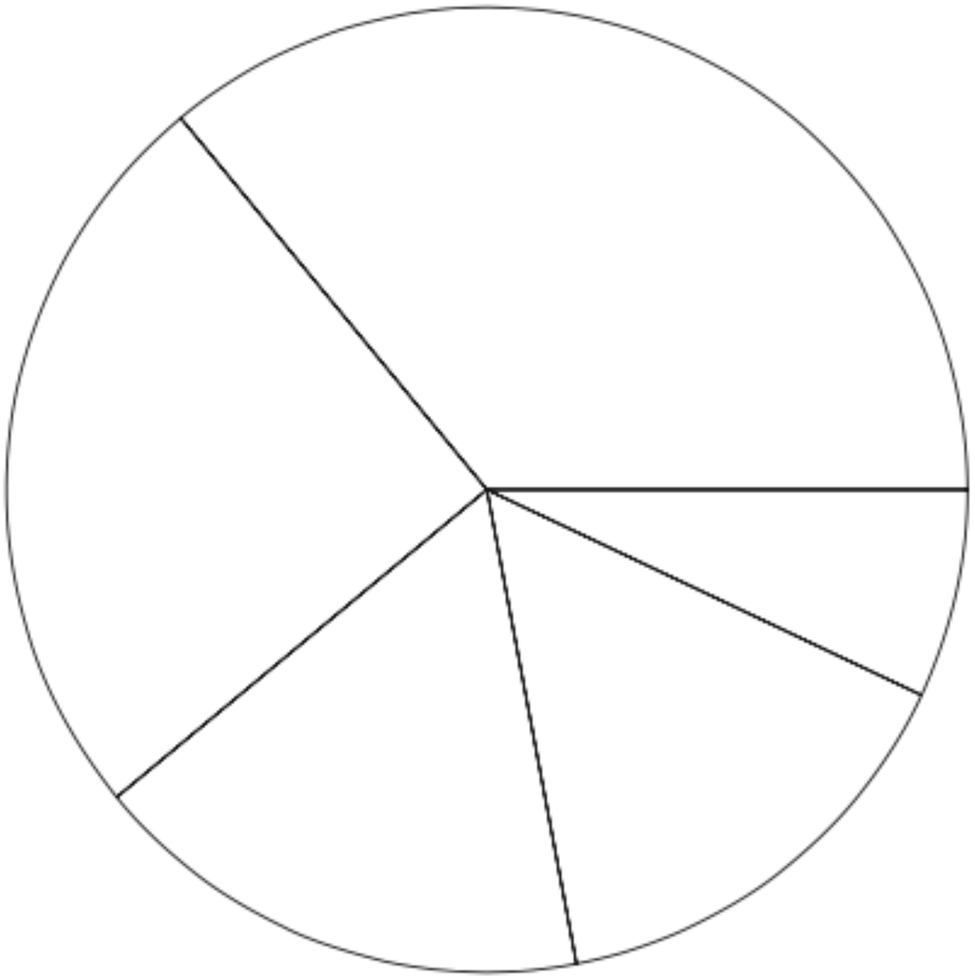
Ponzo Illusion



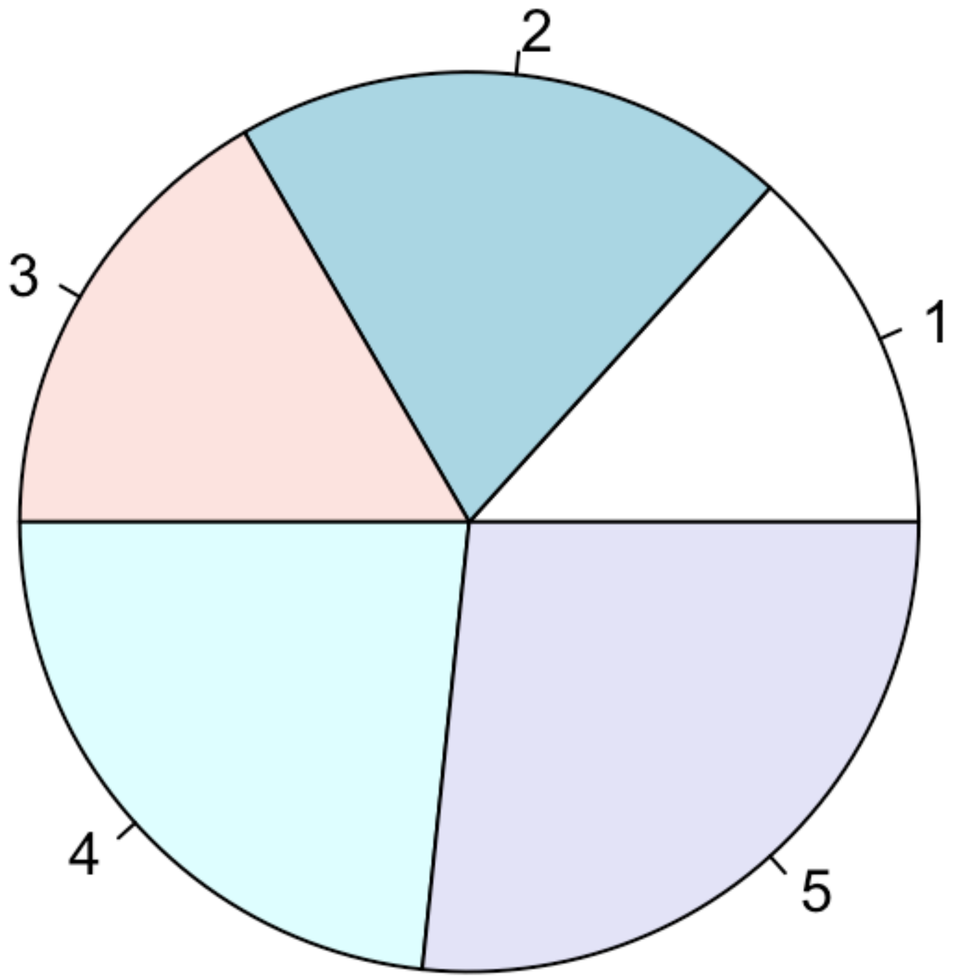
Muller-Lyer Illusion



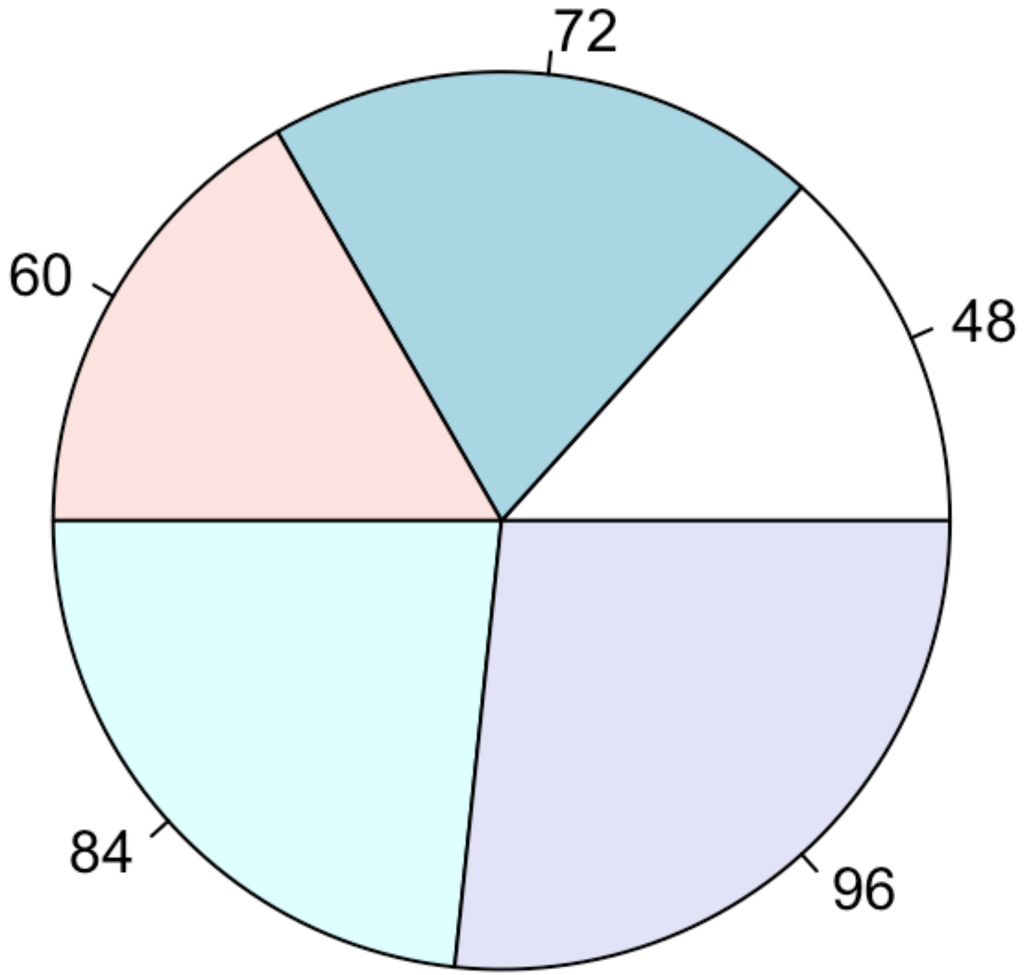
Angle



Angle

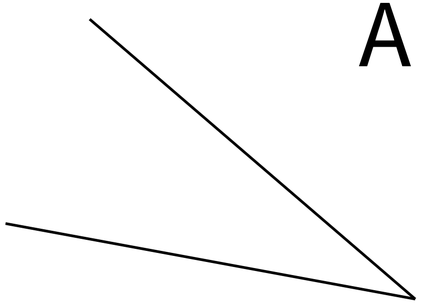


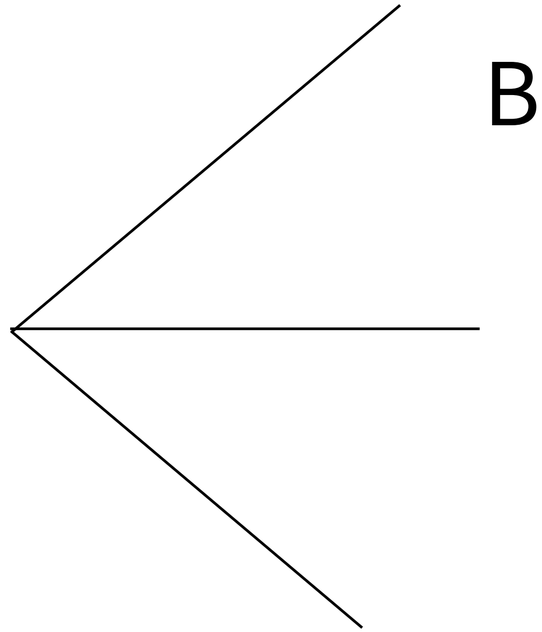
Angle

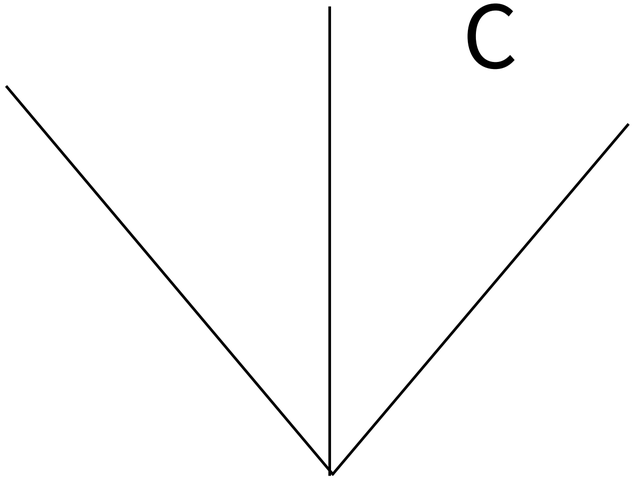


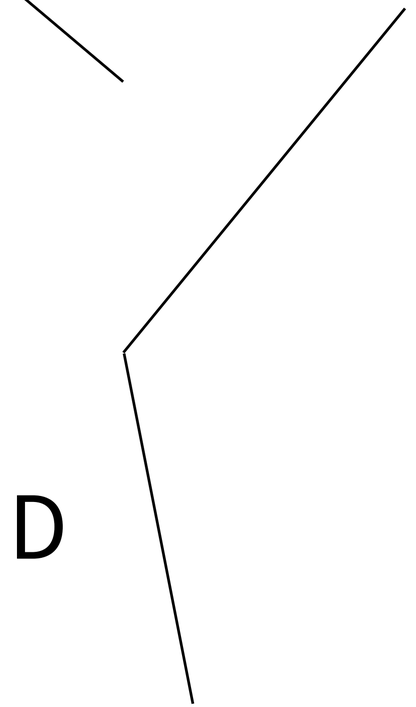
angle measures
in degrees

Angle

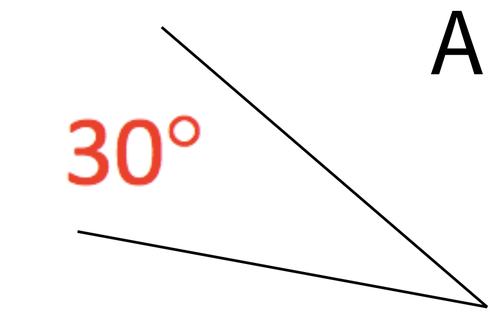


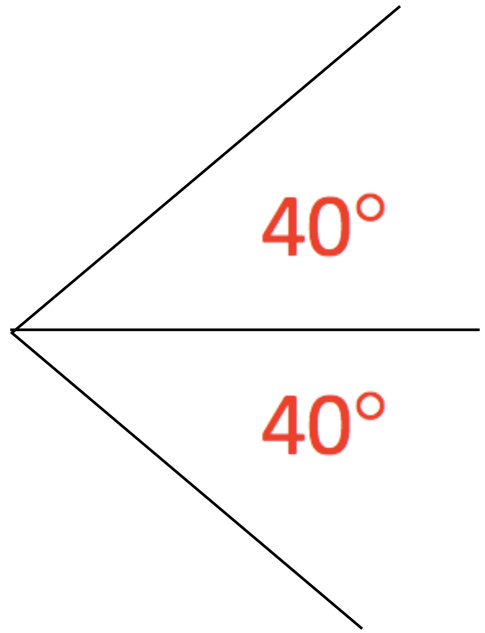


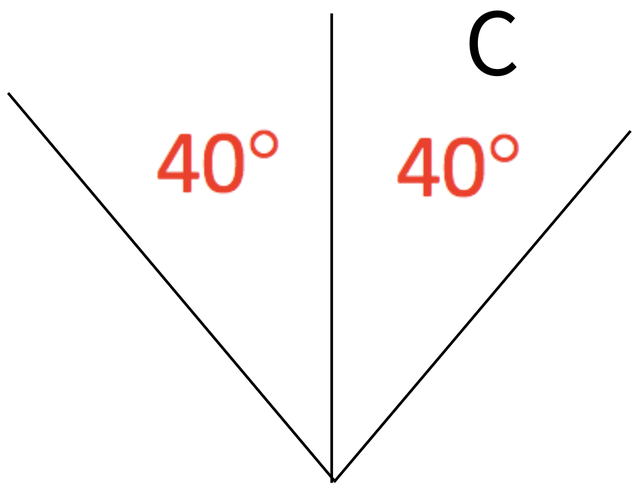


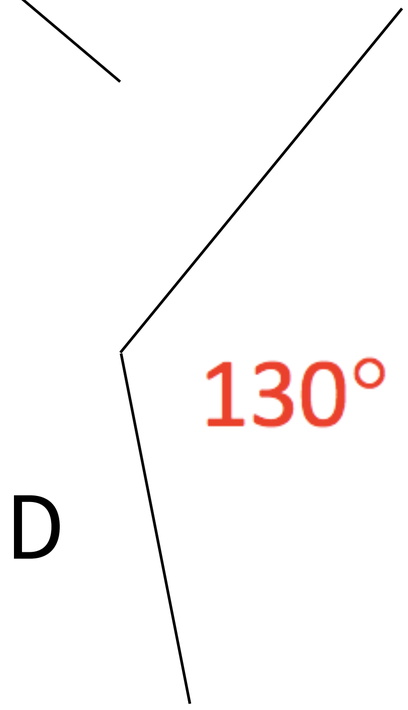


Angle

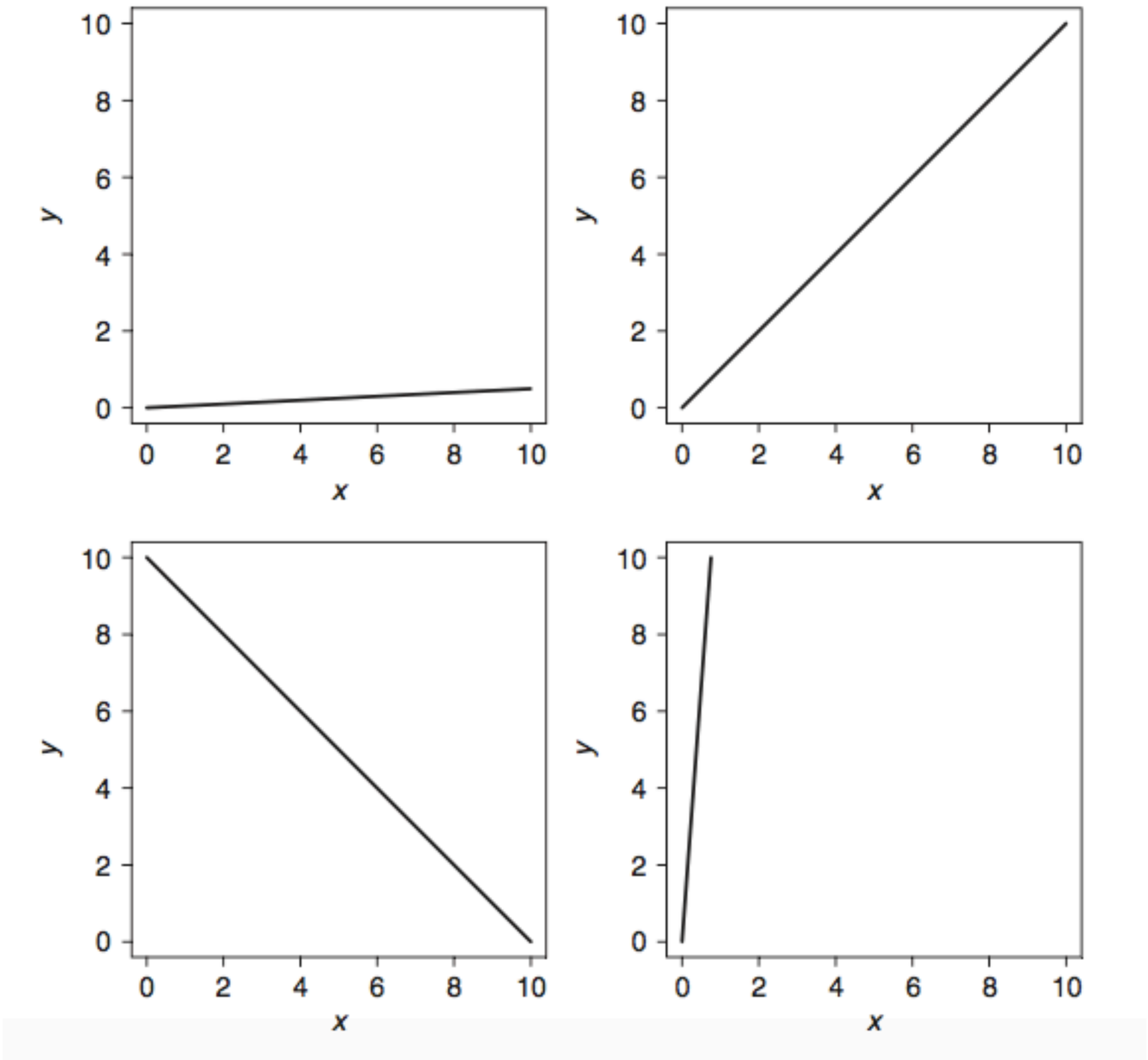




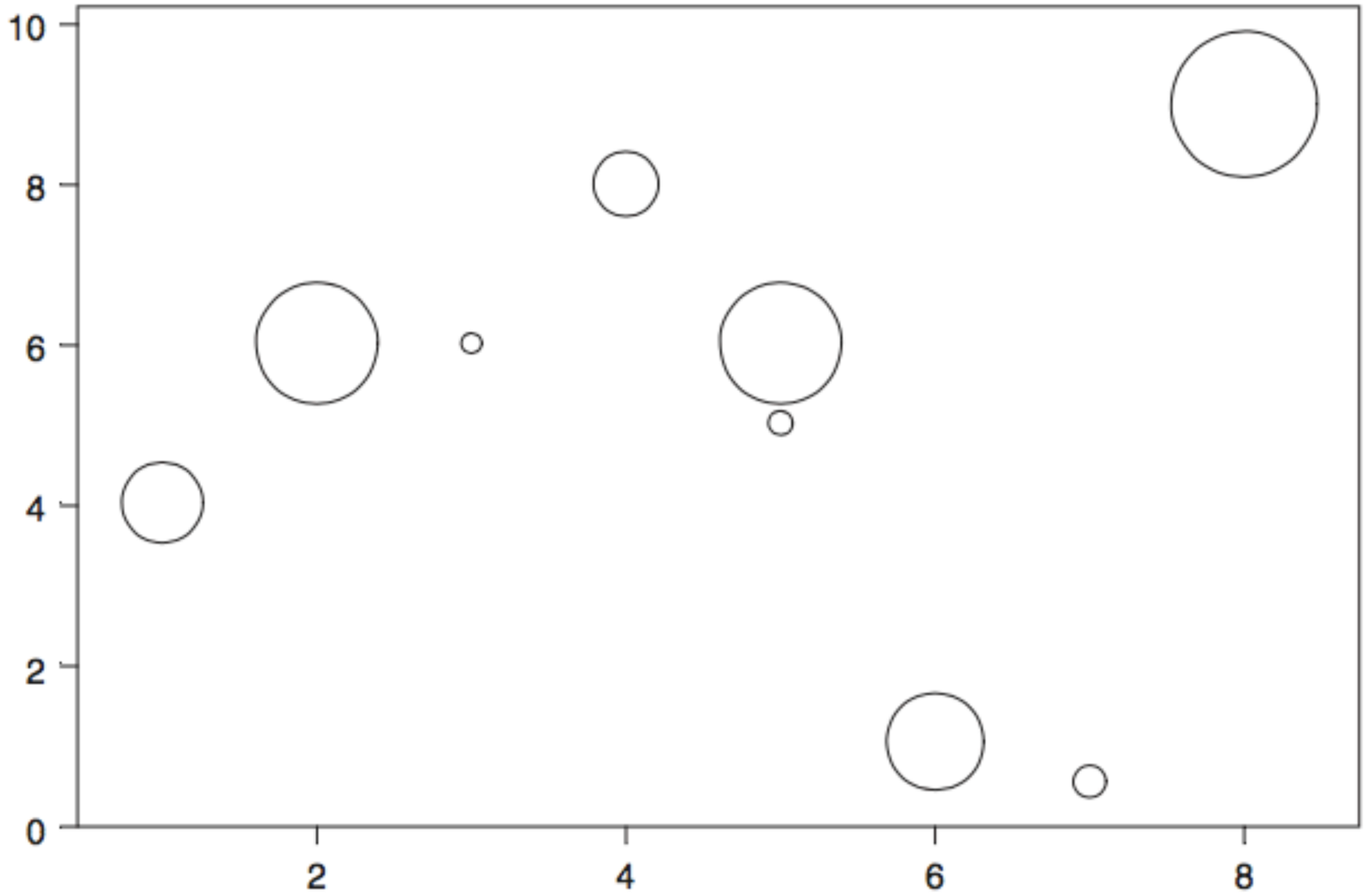




Slope



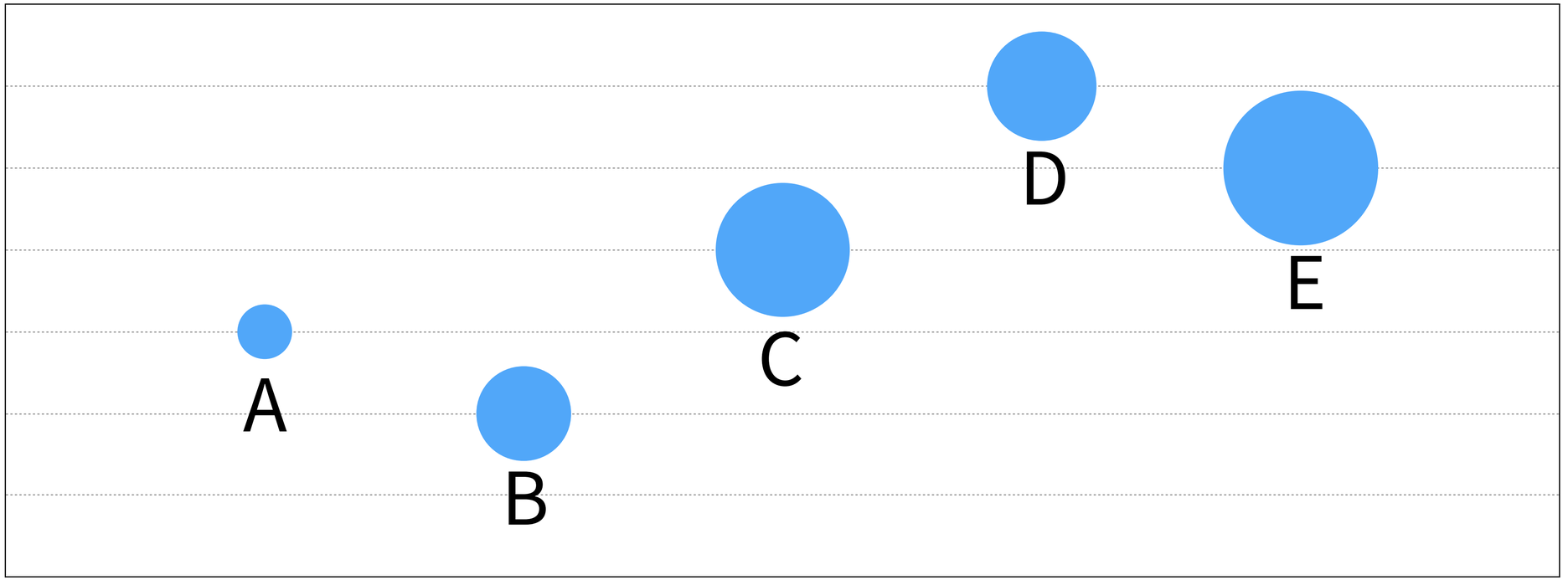
Area



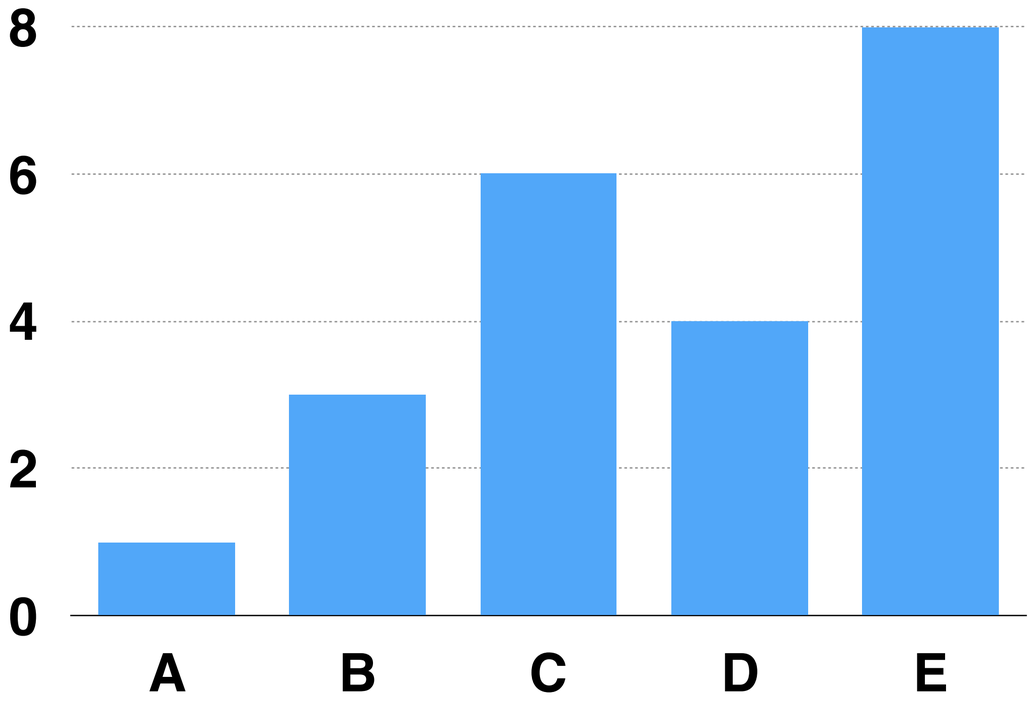
bubble plot

Circles: Area Judgements

If the area of A is 1, what are the areas of the other circles?



Bar Chart (same data)



What is the ratio of the area of a nickel, dime, quarter, and half dollar to the area of a penny?

penny
$0.01

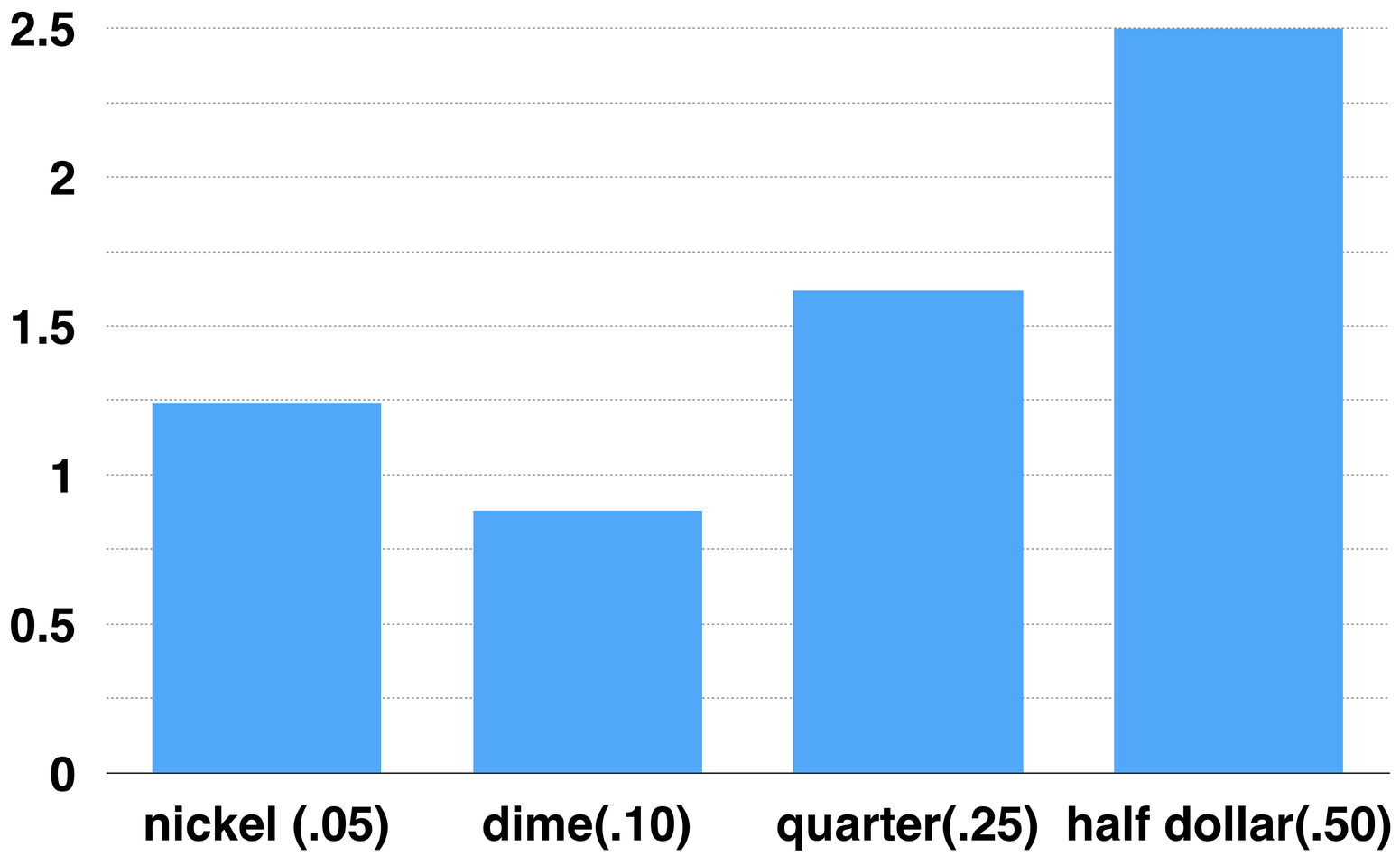
nickel
$0.05

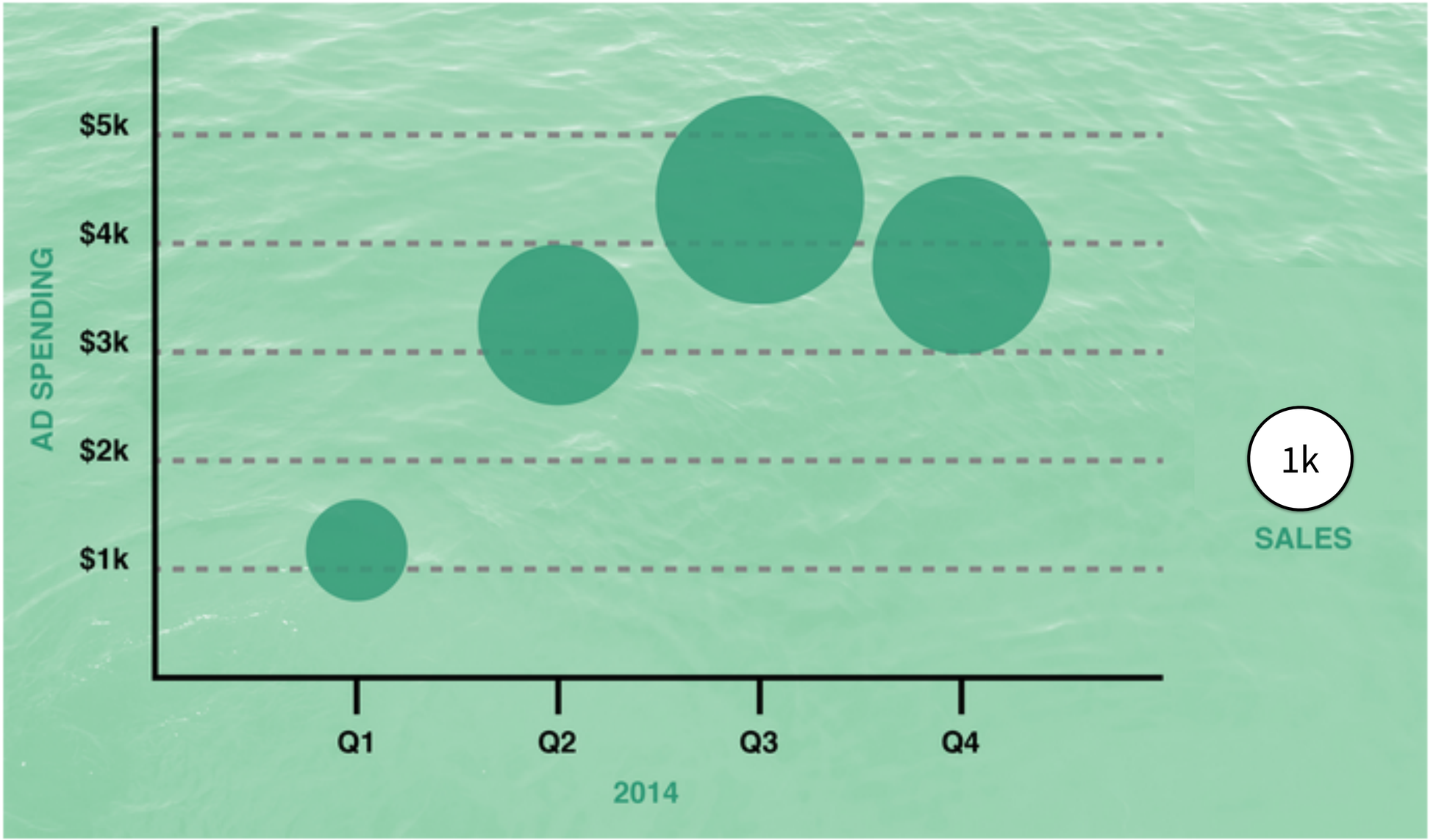
dime
$0.10

quarter
$0.25

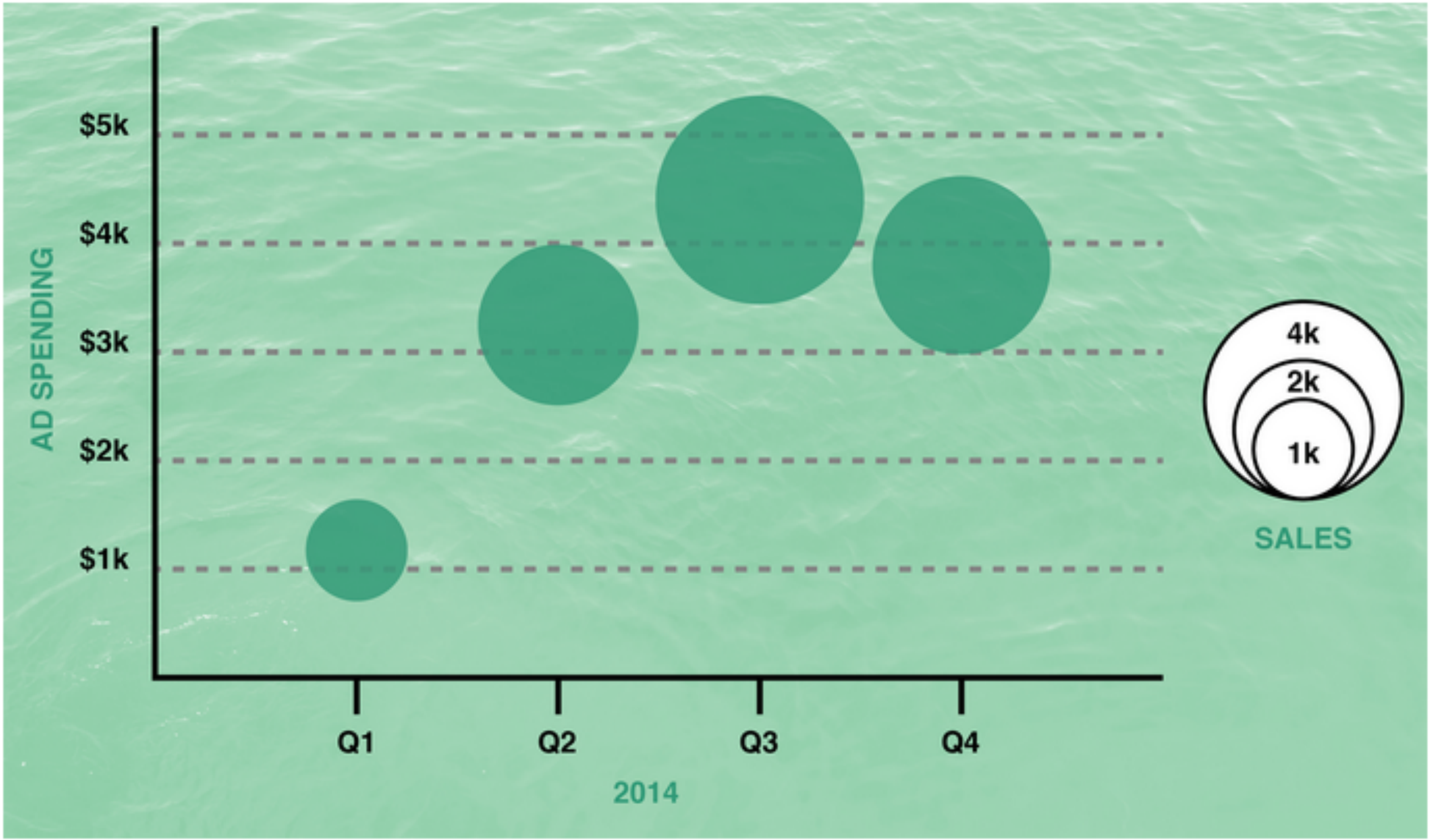
half dollar
$0.50

Ratio of each to the area of a penny

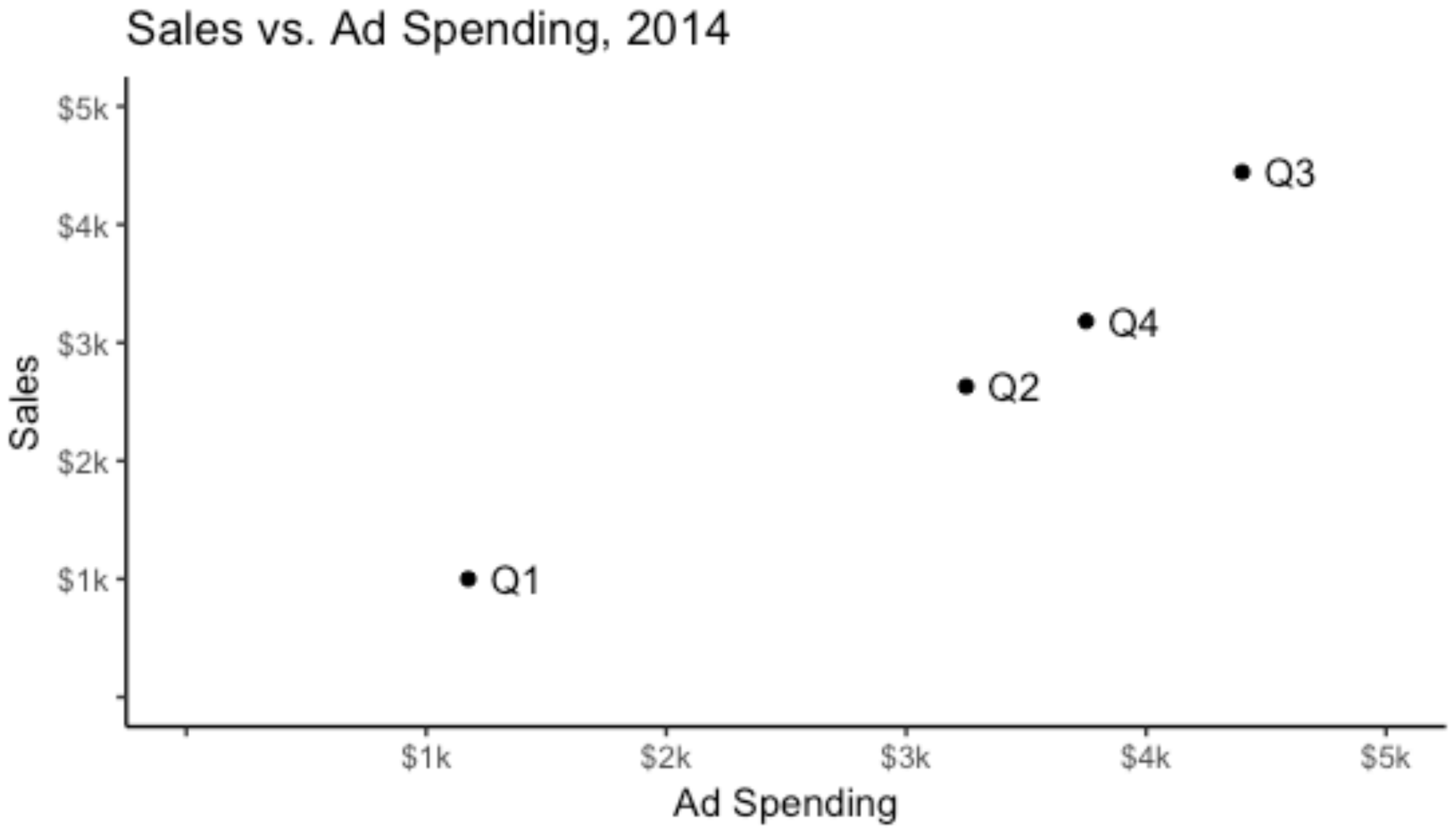




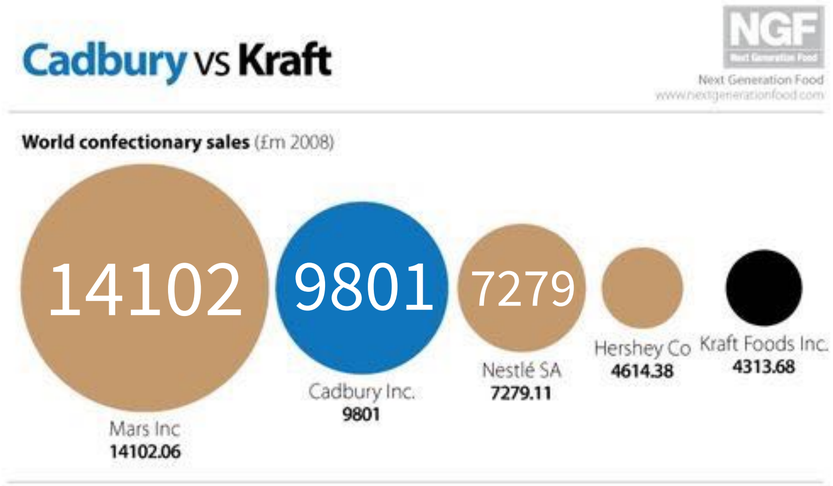
https://visage.co/data-visualization-101-bubble-charts/



https://visage.co/data-visualization-101-bubble-charts/


Sales vs. Ad Spending, 2014
$5k
$4k
$3k
$2k
$1k
Sales
Q3
Q4
Q2
Q1
$1k
$2k
$3k
$4k
$5k
Ad Spending

Area or Diameter?



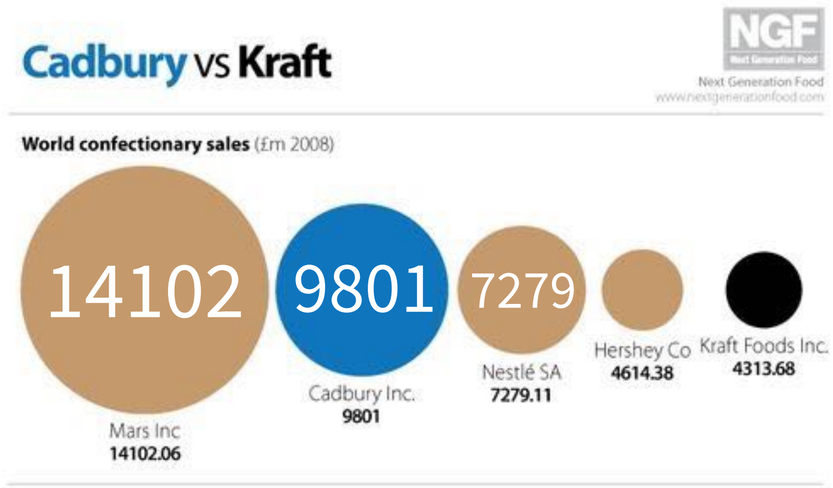
Area of a circle

$$A = \pi r^2$$

ratio of area 1 / area 2 = k

ratio of radius 1 / radius 2 = sqrt(k)

Area or Diameter?



~Diameter

14102/9801 = 1.44

2.85cm/2cm = 1.425

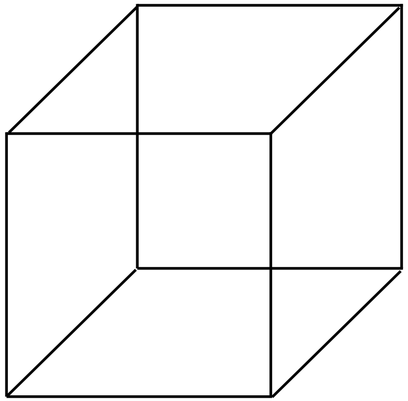
~Area

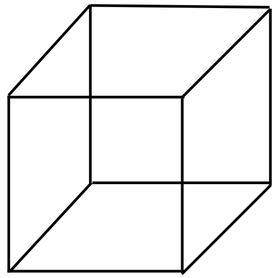
29/14 = 2.07

2.65cm/2cm = 1.32

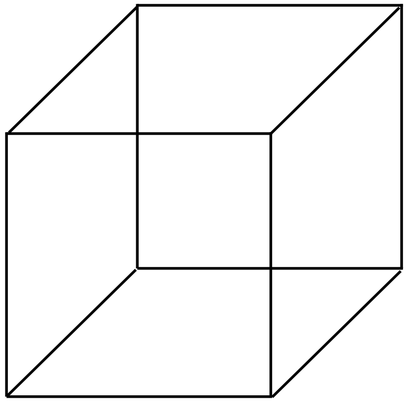
sqrt(2.07) = 1.44

Volume



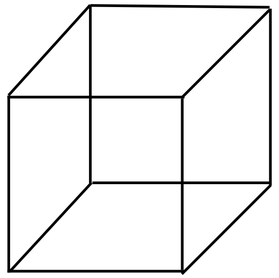


Volume



side = 2.75

volume = 20.7



side = 1.83

volume = 6.13

left cube is
3.38X
larger
(by volume)

"Stevens' Law"

x = actual value perceived value $= Cx^{\beta}$

β = .9 το 1.1 for length

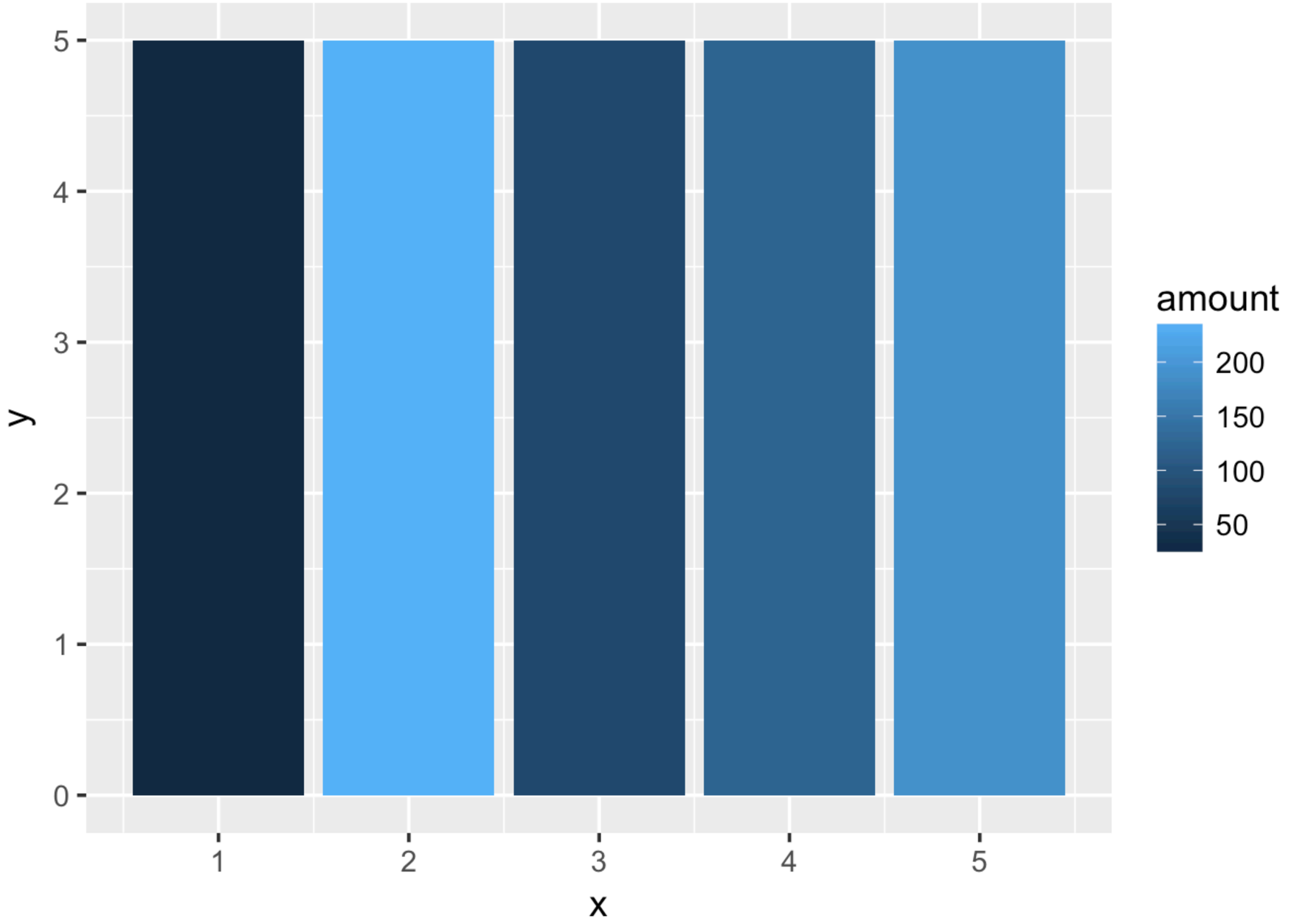
β = .6 to .9 for area

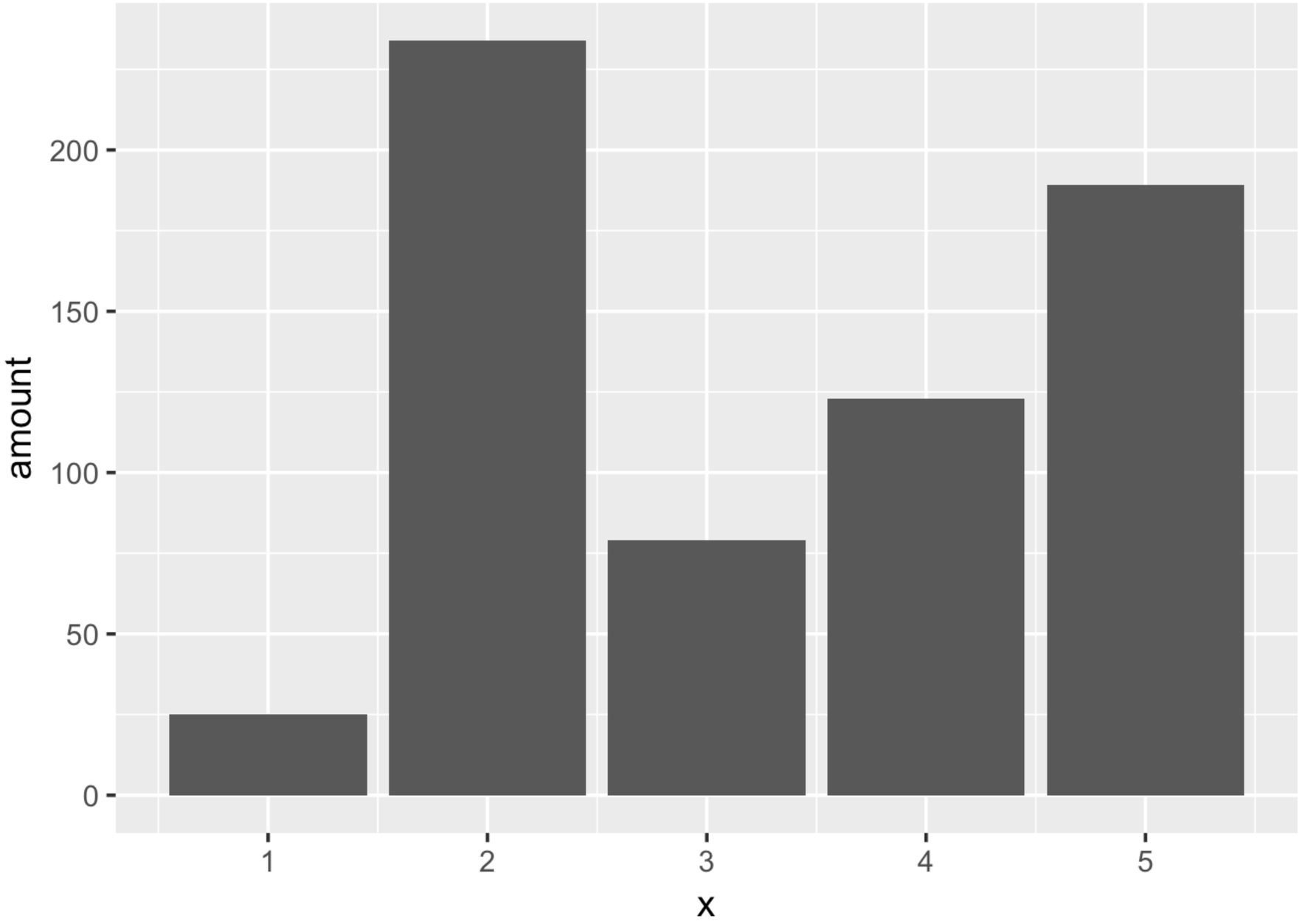
β = .5 το .8 for volume

Color Hue, Saturation Density

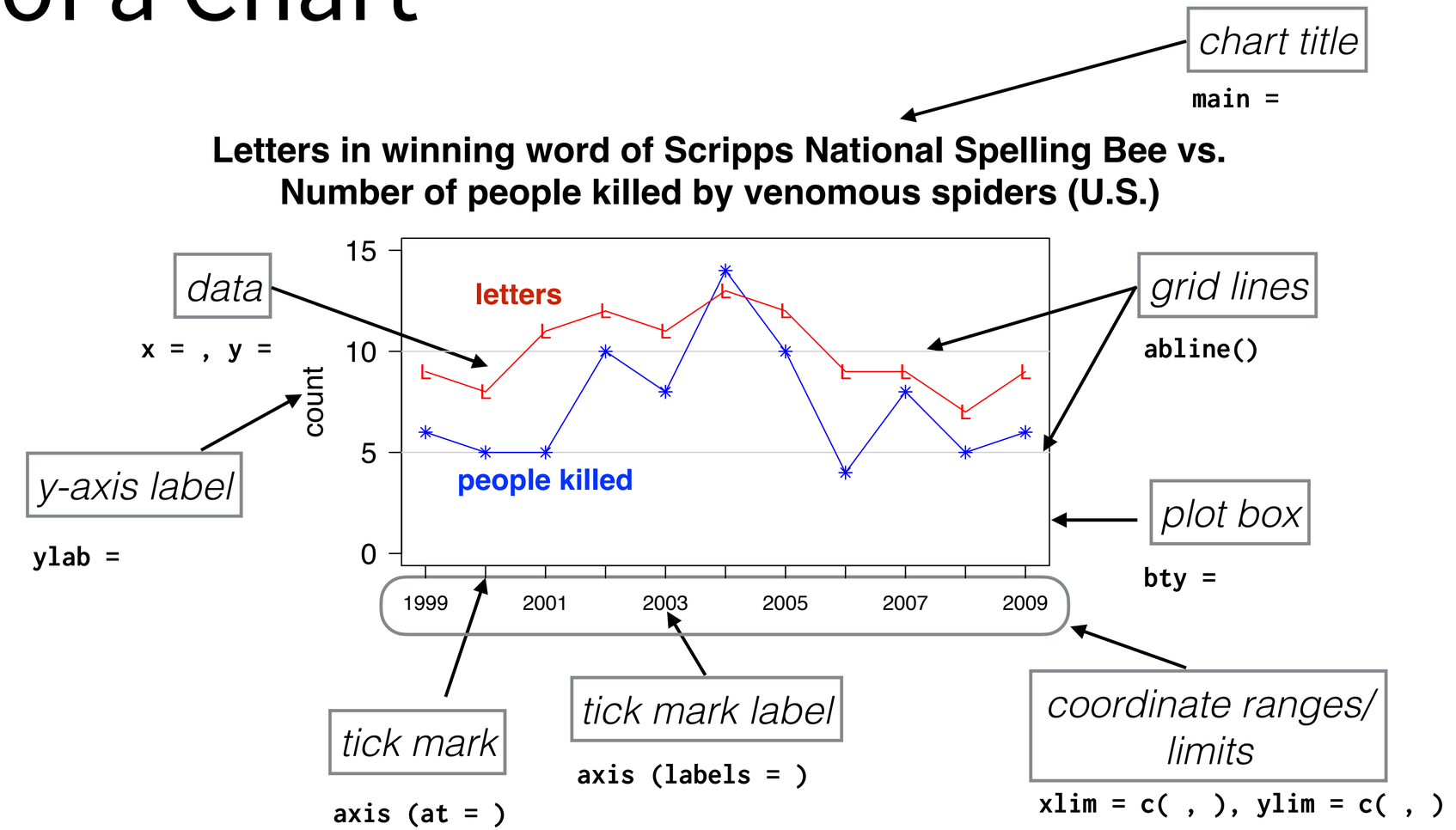
- Hue - 
- Saturation - 
- Lightness - 

Color hue, color saturation, and lightness:
Very effective for categorical variables
NOT for displaying quantitative values


5
4
3
2
1
0
y
1
2
3
4
5
x
amount
200
150
100
50


amount
200
150
100
50
0
1
2
3
4
5
x

Parts of a Chart



R Base Graphics Cheatsheet

SET GRAPHICAL PARAMETERS

*the following can **only** be set with* **`par()`**

par (...)

multiple plots	`mfcol = c(`*nrow,ncol*`)` `mfrow = c(`*nrow,ncol*`)`	*plot margins (outer)*	`oma = c(`*bottom, left, top, right*`)` default: c(0, 0, 0, 0) lines
plot margins	`mar = c(`*bottom, left, top, right*`)` default: c(5.1, 4.1, 4.1, 2.1) lines	*query x & y limits*	`par ("usr")`

CREATE A NEW PLOT

Bar charts **barplot**(*height*, …)
- *bar labels* names.arg =
- *border* border =
- *fill color* col =
- *horizontal* horiz = TRUE

Box plots **boxplot**(*x*, …)
- *horizontal* horizontal = TRUE
- *box labels* names =

Dot plots **dotchart**(*x*, …)
- *dot labels* labels =

Histograms **hist**(*x*, …)
- *breakpts* breaks =

Line charts **plot**(*x*, type = "l")
- *line type* lty = "blank"|0, "solid"|1, "dashed"|2, "dotted"|3
- *line width* lwd =

Scatterplots **plot**(*x*, …)
- *symbol* pch =

REMOVE

axis labels	`ann = FALSE`
axis, tickmarks, and labels	`xaxt = "n"` `yaxt = "n"`
plot box	`bty = "n"`

NOTE: Many of the parameters here can be also be set in `par()`. *See R help for more options.*

ADJUST

allow plotting out of plot region	`xpd = TRUE`
aspect ratio	`asp =`
axis limits	`xlim =, ylim =`
axis lines to match axis limits	`xaxs = "i"` , `yaxs = "i"` *(internal axis calculation)*

ADD TEXT

location

axis labels	`xlab =, ylab =`
subtitle	`sub =`
title	`main =`

style

font face	`font = 1` *(plain)* 2 (*bold)* 3 (*italic)* 4 (*bold italic)*
font family	`family = "serif"` `"sans" "mono"`

size

(magnification factor)

all elements	`cex =`
axis labels	`cex.lab =`
subtitle	`cex.sub =`
tick mark labels	`cex.axis =`
title	`cex.main =`

position

text direction	`las = 1` (horizontal)
justification	`adj = 0 .5 1` (left, center, right)

ADD TO AN EXISTING PLOT

Add new plot [*any plot function*]
(…, add = TRUE)
ex. barplot(x, add = TRUE)

Axes **axis** (*side,…*)
location side = 1 2 3 4
(bottom, left, top, right)
tick mark:
labels labels =
location at =
remove tick = FALSE
rotate text las = 1 (horizontal)

Axis labels **mtext** (*text,…*)
location side = 1 2 3 4
(bottom, left, top, right)
lines to skip line = *(from plot region, default = 0)*
position at = *x or y-coord*
(depending on side)
justification adj = 0 .5 1
(left, center, right)

Lines **lines** (*x,…*)
line style lty =
line width lwd =
color col =

Points **points** (*x,…*)
symbol pch =

color col =
fill color bg = (pch: 21-25 only)

Text **text** (*x, y, text,…*)
position (rel. to x,y) pos = 1 2 3 4
(below, left, above, right)
(default=center)

Title **title** (*main,…*)
axis labels xlab =, ylab =
subtitle sub =
title main =

Cleveland dot plot

```
source("dotchartsolid.R")
data <- read.csv("countries2012.csv")
index <- seq(from = 1,to = 179, by = 4)
sample <- data$TFR[index]
names(sample) <- data$COUNTRY[index]
sample <- sample[order(sample)]
par(mar = c(5, 10, 4, 2))
dotchartsolid(sample, cex = .8, pch = 16, xlim = c(1,7),
              main = "Total Fertility Rate by Country",
              xlab = "average births per woman",
              adj = 1, cex.main = 2,cex.lab = 1.5)
abline (v=2, col = "red")
text (2, 12.5, "replacement rate", cex = .7, pos = 4,
      col = "red")
```