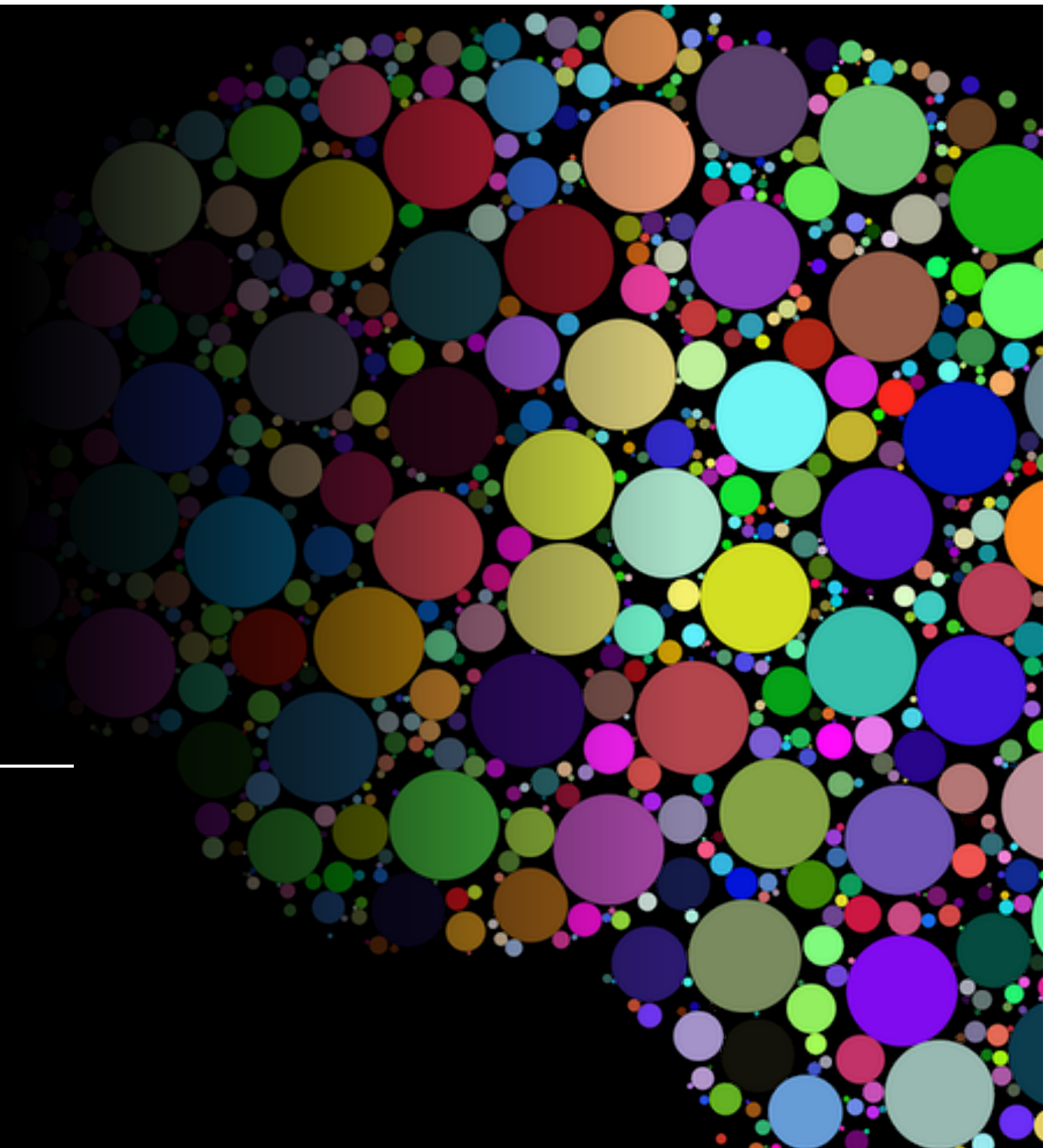




# Visual Displays of Data

---

Chapter 3



# Uses of Graphs

- Positive and negative uses
  - Can accurately and succinctly present information
  - Can reveal/conceal complicated data
- Graphing in the information age: a critical skill
  - <http://www.datavis.ca/gallery/>

# “The Most Misleading Graph Ever Published”

## How to Mislead with Graphs

The Cost and Quality of Higher Education



# “The Most Misleading Graph Ever Published”

- So why bad?
  - The times are not the same for both lines (going across)
  - The scales are not the same
    - So the starting point is also a problem
  - Rankings ... down = BETTER, so why did they graph ranks that way?

# Techniques for Misleading

- The false face validity lie
  - Method seems to represent what it says, but does not actually.
    - e.g., using yelling as a measure of aggression

# Techniques for Misleading

- The biased scale lie
  - Scaling to skew the results, or leading questions
  - i.e. goodreads book scale (did not like, it was ok, liked it, really liked it, it was amazing)

# Techniques for Misleading

- The sneaky sample lie
  - When participants are preselected or self selected to provide data.
  - Like rate my professor!

# More Techniques

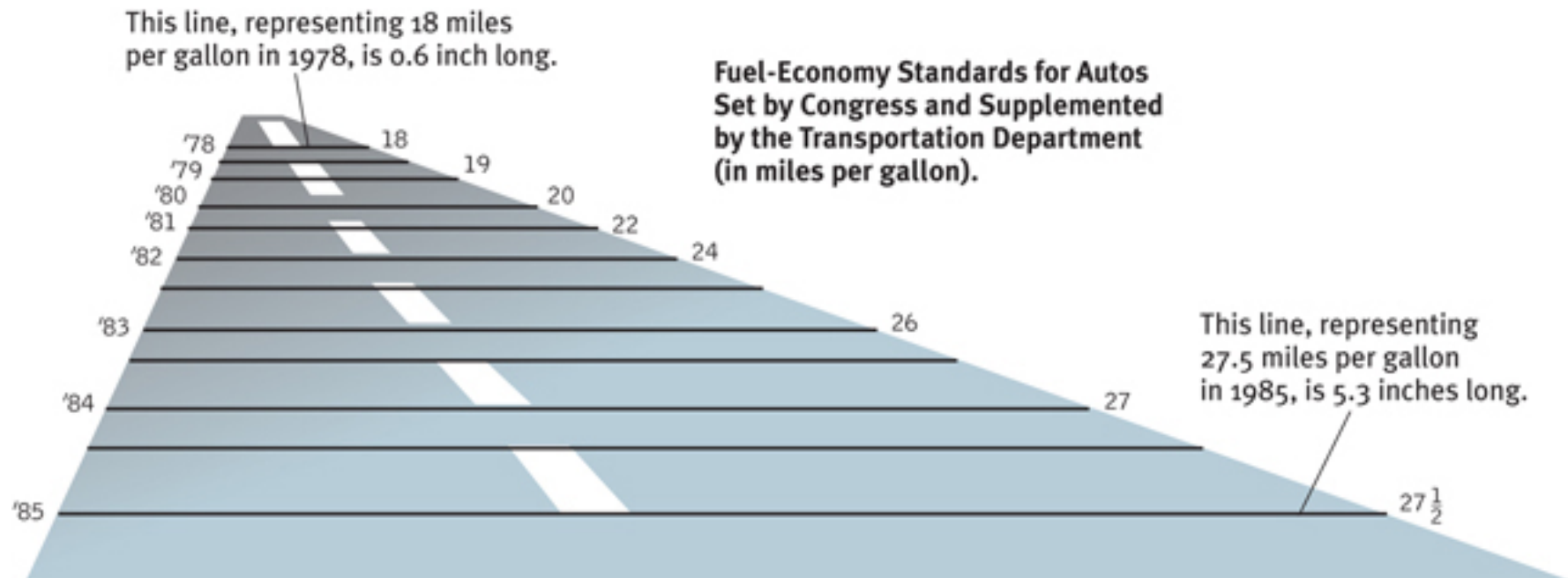
- The interpolation lie
  - Assumes that a value between 2 data points follows the same pattern
- The extrapolation lie
  - Assumes knowledge outside of the study



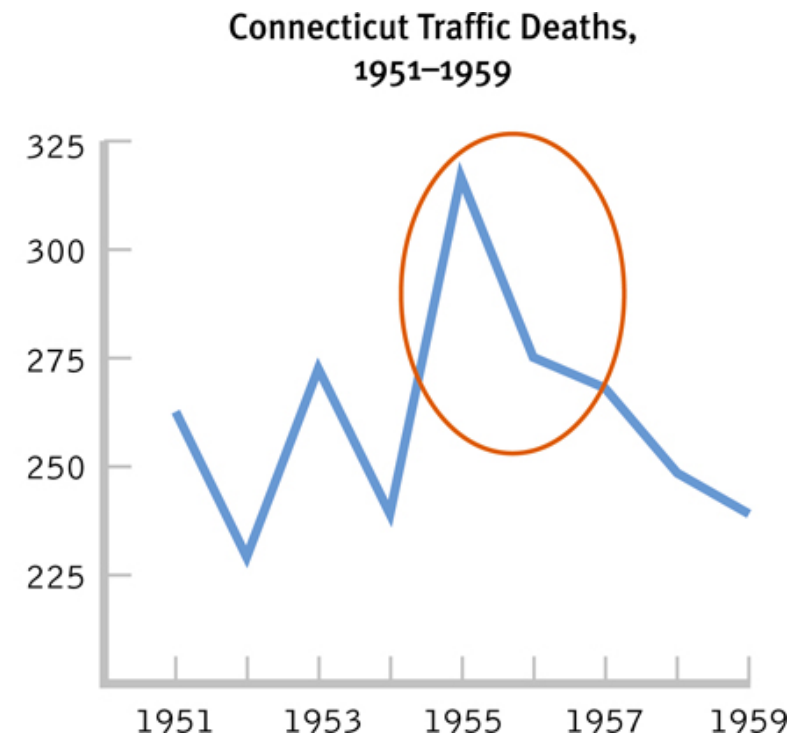
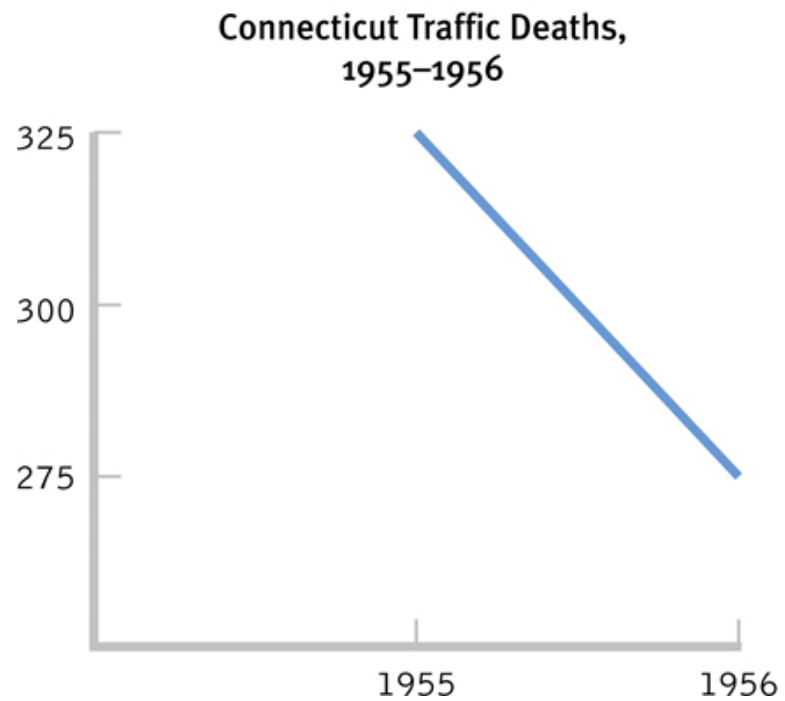
# More Techniques

- The inaccurate values lie
  - Uses scaling to distort portions of the data
- The outright lie:
  - Making up data!

# The Inaccurate Value Lie



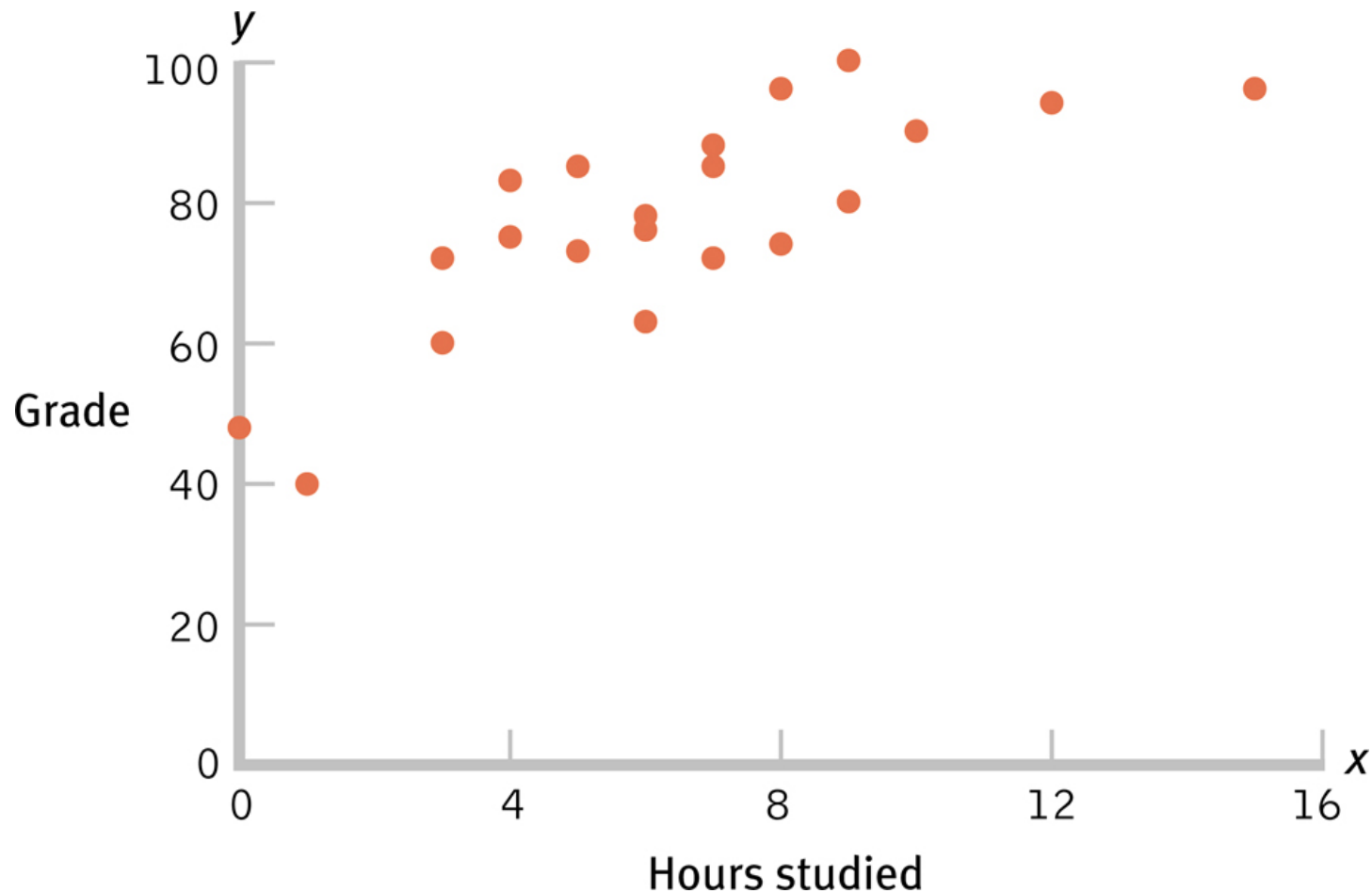
Which of the two graphs is misleading?



# Common Types of Graphs

- Scatterplots
  - Graphs that depict the relation between two scale variables
  - Range-frame (!!)
  - x and y axis show the range of the variable
  - Observing every data point
    - Linear relationships – relationship between variables is a straight line
    - Nonlinear relationships – relationship between variables is curved
      - Performance & Anxiety

# Scatterplot of Hours Studied and Statistics Grade



# How-To R

- Be sure you load the ggplot library!
  - We are going to use the airquality dataset.
  - `data(airquality), head(airquality)`

# How-To R

- First build a blank plot.
- `myplot = ggplot(dataset, aes(X axis column, Y axis column))`
  - Example: `myplot = ggplot(airquality, aes(Temp, Ozone))`

# How-To R

- Add dots to it!
  - `geom_point()`
- Man, that's still an ugly graph...how can we clean it up?



# How-To R

- Run this cleanup code:
- ```
cleanup = theme(panel.grid.major = element_blank(),  
  panel.grid.minor = element_blank(),  
  panel.background = element_blank(),  
  axis.line.x = element_line(color = "black"),  
  axis.line.y = element_line(color = "black"),  
  legend.key = element_rect(fill = "white"),  
  text = element_text(size = 15))
```

# How-To R

- Now do,
- `myplot + geom_point() + cleanup`
  - Much nicer!
- Now, how can we change other things about the graph?
  - How about x and y axis labels?

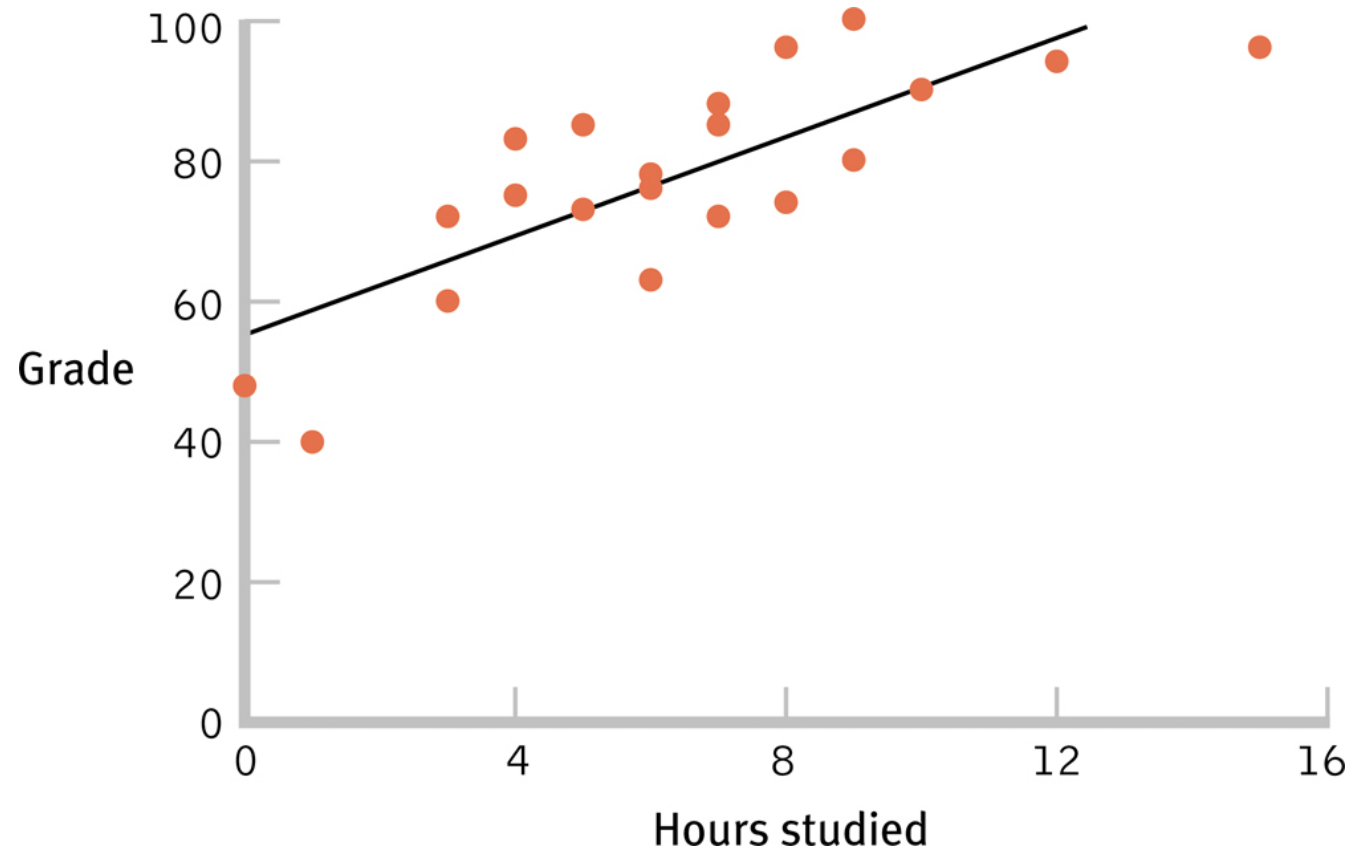
# How-To R

- + xlab(*"Text goes here"*)
- + ylab(*"Text goes here"*)

# Line Graphs

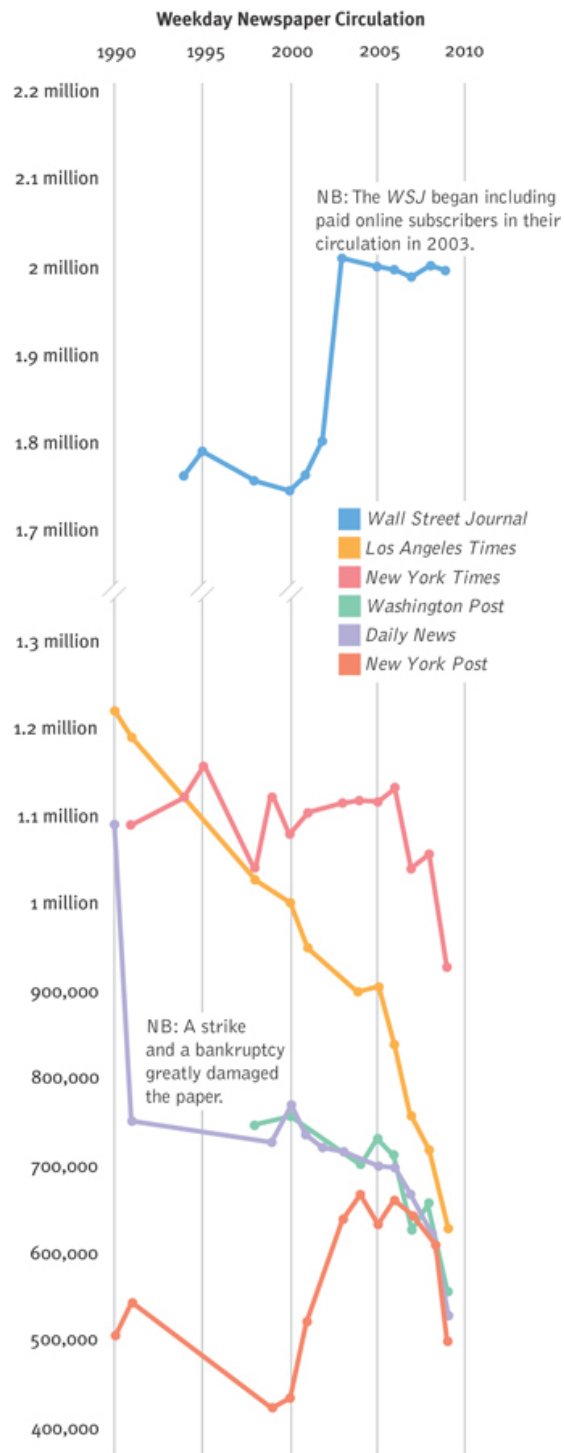
- Line graphs are best for scale variables (remember this is interval/ratio)
  - Especially useful for trends over time
- Line of best fit
- Time series plot

# The Line of Best Fit



# How-To R

- Let's add a line of best fit!
- + `geom_smooth()`
  - Oh, that's not quite straight.
  - Let's do this instead!
  - `geom_smooth(method = "lm")`



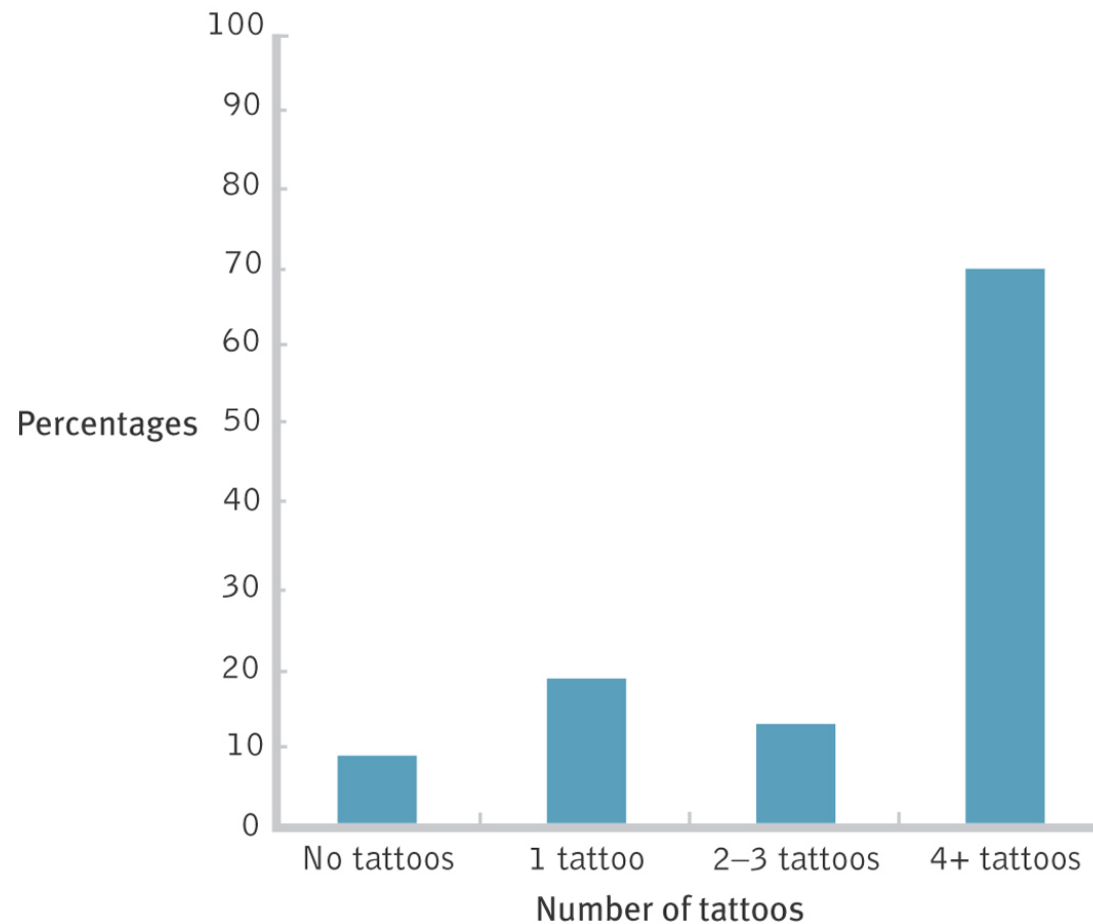
# Weekly Newspaper Circulation

# Bar Graphs

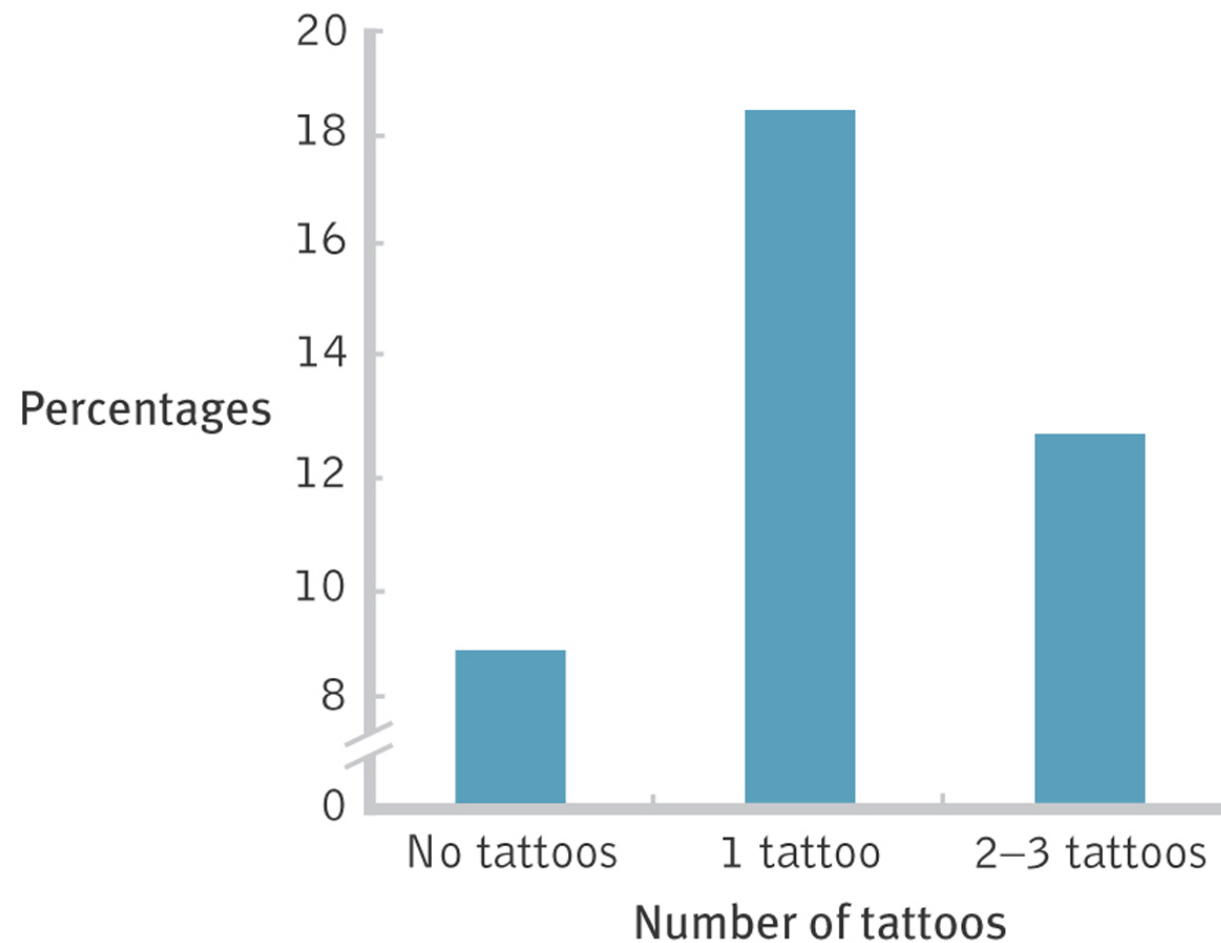
- When the independent variable is nominal or ordinal and the dependent variable is scale
  - Meaning  $X$  = groups,  $Y$  = interval/ratio
  - Each bar usually represents an AVERAGE score.



# Bar Graphs Highlight Differences Between Means



# Deceiving with the Scale



# How-To R

- First, build a blank plot.
  - `myplot = ggplot(dataset, aes(X axis column, Y axis column))`
  - But now, the X axis column will be categorical, the Y axis column will be continuous

# How-To R

- `myplot = ggplot(airquality, aes(Month, Temp))`
  - Month is an integer at the month, let's fix it!
  - The `factor()` command allows us to take numeric variables and switch them to text.
  - `factor(column,`  
    `levels = c(stuff that's in the column already),`  
    `Labels = c(stuff you want to change it to))`

# How-To R

- + stat\_summary(fun.y = mean,  
geom = "bar",  
fill = "White",  
color = "Black")

# How-To R

- + stat\_summary(fun.data = mean\_cl\_normal,  
geom = "errorbar",  
position = position\_dodge(width = 0.90),  
width = 0.2)

# Check Your Learning

- What is the best type of graph to depict the following:
  - Depression levels and stress levels for 150 university students. Is depression related to stress?
  - Mean years of education for six regions of the United States. Are education levels higher in some regions than in others?

# Choosing the Graph Based on Variables

- One scale variable (with frequencies): histogram or frequency polygon
- One scale independent and one scale dependent variable: scatterplot or line graph
- One nominal or ordinal independent and one scale dependent variable: bar graph
- Two+ nominal or ordinal independent and one interval dependent variable: bar graph