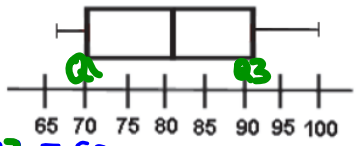
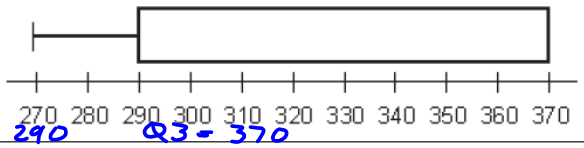
HW 4-1

Data	Lower outlier fence	Upper outlier fence	Example outlier values
18, 24, 32, 34, 45, 55, 78 Q1 = 24 Q3 = 55 IQR = 46.5	-22.5	101.5	110
14, 13, 4, 14, 10, 2, 13, 5, 0, 0, 13, 3, 9, 2, 10, 11, 13, 5, 14 Q1 = 3 Q3 = 13 IQR = 15	-12	28	30
 Q1 = 70 Q3 = 92 IQR = 33	~~28.5~~ 37	~~122.5~~ 125	30 or 130
 Q1 = 290 Q3 = 370 IQR = 120	170	490	150 or 500

3. Why are quartiles generally more reliable for judging outliers than mean and standard deviation?

   Means and standard deviations are not resistant to the effects of outliers, while quartiles are resistant.

4. What are two things we should never do with outliers?

   We should never ignore outliers or throw them away.

5. What should we always do before excluding outliers in data analysis?

   We should always try to gather more information and try to find out why the value is unusual. If the value is an error, we can correct it. Finally, we should perform the analysis both with and without the outlier.

From the text #6, 10

6. Still rockin'.

a) The histogram and boxplot of the distribution of "crowd crush" victims' ages both show that a typical crowd crush victim was approximately 18 - 20 years of age, that the range of ages is 36 years, that there are two outliers, one victim at age 36 - 38 and another victim at age 46 – 48.

b) This histogram shows that there may have been two modes in the distribution of ages of "crowd crush" victims, one at 18 - 20 years of age and another at 22 – 24 years of age. Boxplots, in general, can show symmetry and skewness, but not features of shape like bimodality or uniformity.

c) Median is the better measure of center, since the distribution of ages has outliers. Median is more resistant to outliers than the mean.

d) IQR is a better measure of spread, since the distribution of ages has outliers. IQR is more resistant to outliers than the standard deviation.

10. Graduation?

Upper Fence: $Q3 + 1.5(IQR) = 74.75 + 1.5(74.75 - 59.15)$
$= 74.75 + 23.4$
$= 98.15$

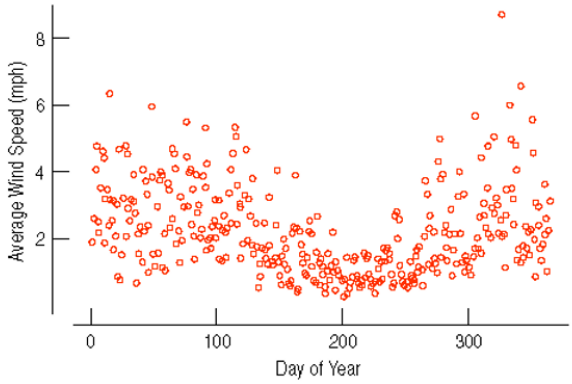
Lower Fence: $Q1 - 1.5(IQR) = 59.15 - 1.5(74.75 - 59.15)$
$= 59.15 - 23.4$
$= 35.75$

Since the maximum value of the distribution of the percent of incoming freshmen who graduate on time is 87.4% and the upper fence is 98.15%, there are no high outliers. Likewise, since the minimum is 43.2% and the lower fence is 35.75%, there are no low outliers. Since the minimum and maximum percentages are within the fences, all percentages must be within the fences.

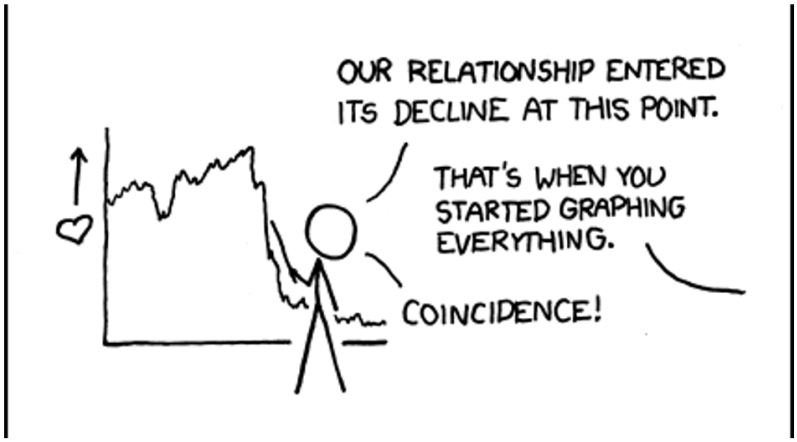
Timeplots and Shifting/Scaling Data

Timeplots: Order, Please!

- For some data sets, we are interested in how the data behave over time. In these cases, we construct timeplots of the data.

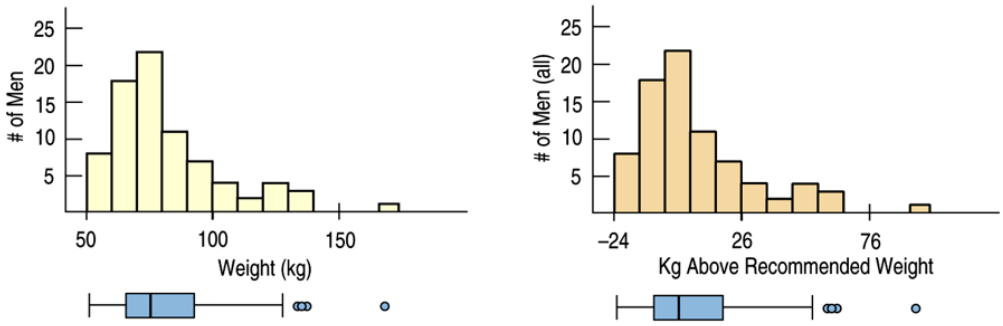


A time plot of interest



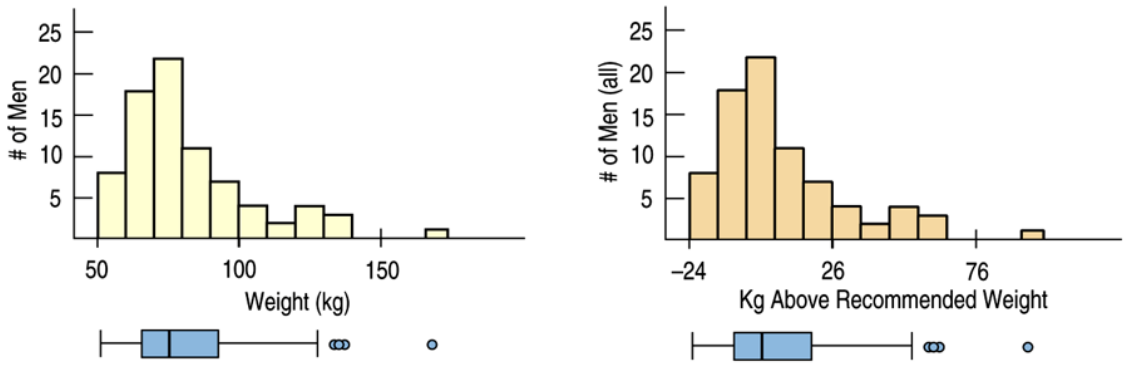
Shifting Data (cont.)

- The following histograms show a shift from men's actual weights to kilograms above recommended weight (by subtracting 74 kg from each):



Shifting Data (cont.)

- What is center of the two distributions?
- What is the IQR of the two distributions?



Summary of effects of shifting:

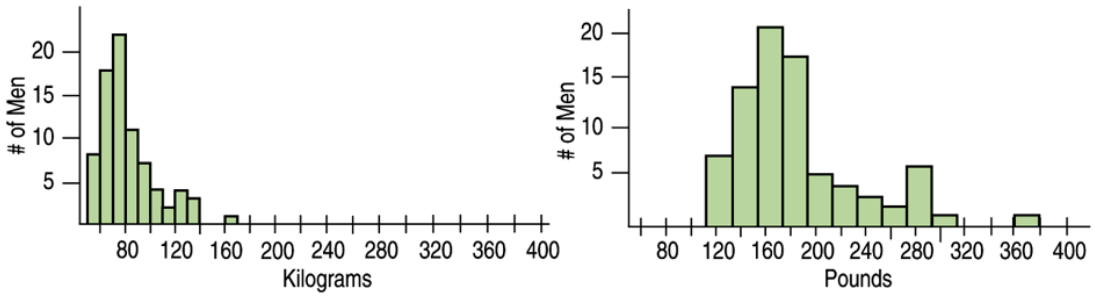
Shape: Stays exactly the same

Center: Shifts ↓ 74 kg. (subtract 74 kg)

Spread: Stays the same

Rescaling Data (cont.)

- The men's weight data set measured weights in kilograms. If we want to think about these weights in pounds, we would rescale the data (multiply by 2.2):



Rescaling Data

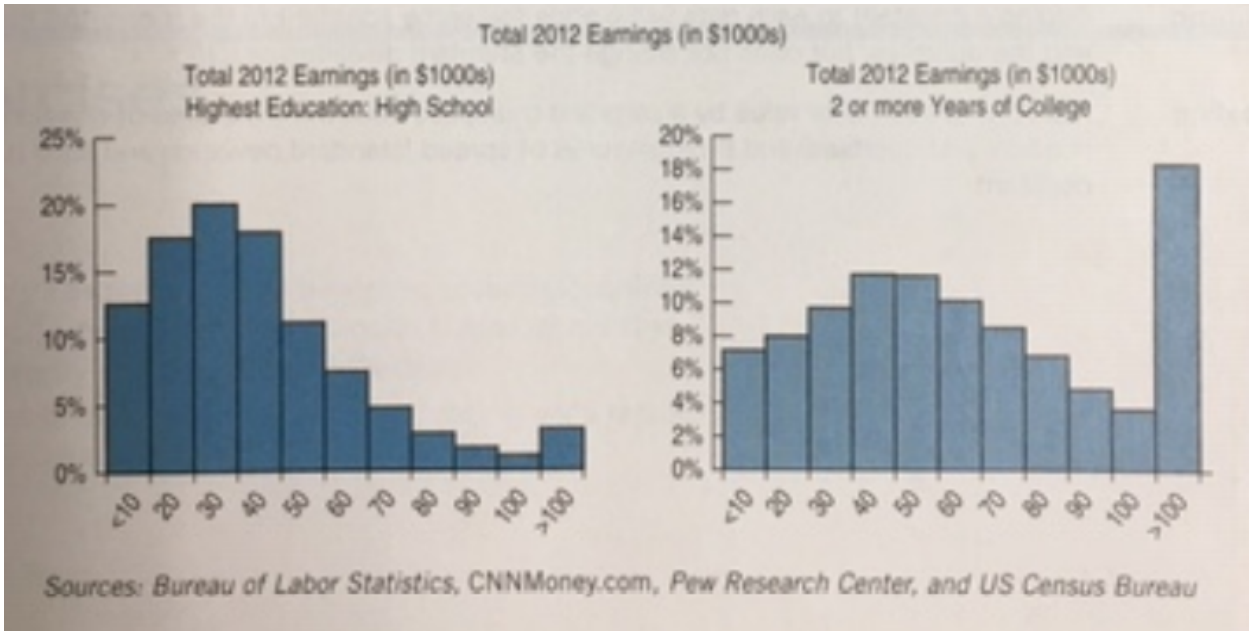
- Rescaling data:
  - When we divide or multiply all the data values by any constant value, all measures of position (such as the mean, median and percentiles) and measures of spread (such as the range, IQR, and standard deviation) are divided or multiplied by that same constant value.

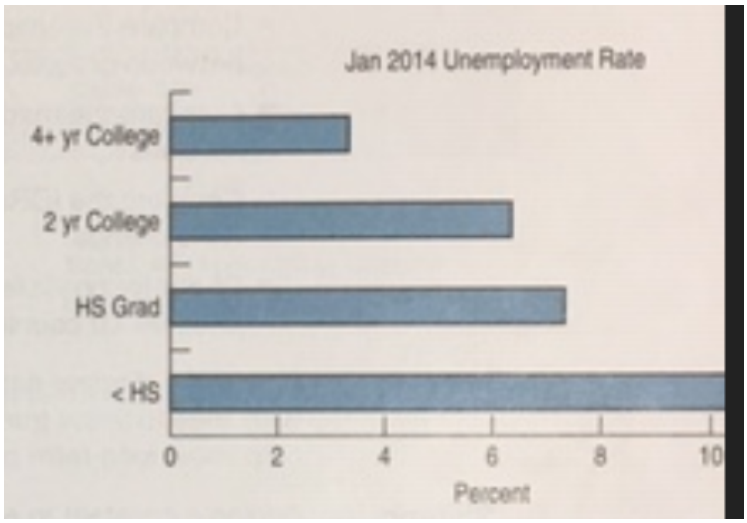
Suppose the class took a 40-point quiz. Results show a mean score of 30, median 32, IQR 8, SD 6, min 12, and Q1 27. (Suppose YOU got a 35.)

What happens to each of the statistics if…

- I decide to weight the quiz as 50 points, and will add 10 points to every score. Your score is now 45.
- I decide to weight the quiz as 80 points, and double each score. Your score is now 70.
- I decide to count the quiz as 100 points; I'll double each score and add 20 points. Your score is now 90.

		×2
min	12	24
max	40	80
Range	28	56


Total 2012 Earnings (in $1000s)
Total 2012 Earnings (in $1000s)
Highest Education: High School
25%
20%
15%
10%
5%
0%
<10
20
30
40
50
60
70
80
90
100
>100
Total 2012 Earnings (in $1000s)
2 or more Years of College
20%
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
<10
20
30
40
50
60
70
80
90
100
>100
Sources: Bureau of Labor Statistics, CNNMoney.com, Pew Research Center, and US Census Bureau


Jan 2014 Unemployment Rate
4+ yr College
2 yr College
HS Grad
< HS
0
2
4
6
8
10
Percent

Homework:

Worksheet - Practice Shifting and Scaling

Read page 94-98 (top)

pg 105 # 11,14,16