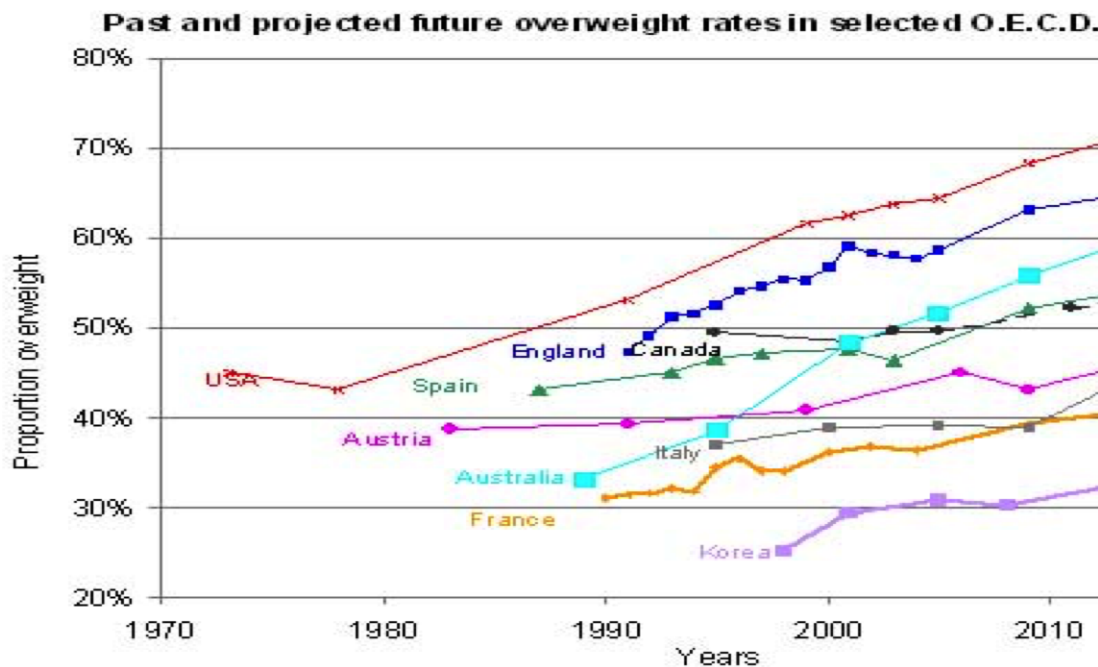


City Semester
 Problem Set #3
Show all work for full credit.

- Calculate the following summary statistics for the given data set: $\{1, 2, 2, 3, 3, 5, 9, 11\}$
 - Find \bar{X} .
 - Find the variance σ^2 .
 - Find the standard deviation σ .
 - Explain how why σ is used as a measure of spread.
 - Give the five number summary for this data set.
- A few years ago then mayor Bloomberg made it a personal quest to control what New Yorkers eat in an effort to improve health largely motivated by statistics like this: the percentage of New York State adults who are overweight or obese increased from 42% in 1997 to 60% in 2008. (source: NYS Dept of Health)

New Yorkers compare slightly more favorably to US citizens as a whole, but not much. The graph below shows the percent of the population in various countries that are overweight.



- According to the graph, what was the approximate percentage of the US population that was overweight in 1995? Is your answer an interpolation or an extrapolation?
- The graph is cut off at about 2013. Based on the portion of the graph that is shown, about what percent of the US population do you think will be overweight in 2020? Is your answer an interpolation or an extrapolation?

- (c) If you are confident this relationship is linear what portion of the US population do you think will be overweight in 2100? How did you arrive at that answer? Are you more or less confident in this answer than in part b?
3. 🟡 Average Hourly Wage for workers in NYC 2009 to 2013
- (a) Open the Fathom data file AvgHourlyEarnings-NYC and fit a linear model to the data with Months since Jan09 as the input and AverageHourlyEarning as the output.
 - (b) Right-click on the graph after youve plotted your function and choose Make Residual Plot (near the bottom). What is the residual plot showing you?
 - (c) Looking at the residual plot, do you think a linear model is appropriate for this data? Do you think your model is a useful predictor of data values? Explain.
4. 🟡 Revisit Average Hourly Wage in NYC 2009 to 2013
- (a) Open data file AvgHourlyEarnings-NYC and fully analyze the data according to the 5 steps shown on the slides (you can use your model from the last problem set if you still have it).
 - (b) Write a sentence that explains the meaning of the slope of your linear model within the context.
 - (c) Use the model to predict what the Average Hourly Wage would be for Jan 2015 (72 months after Jan 09)
 - (d) According to the slope of the model, what was the average yearly change in hourly wage?
5. 🟡 Lastly, let's look at obesity rates in NY state. Keep in mind this is obesity as measured by BMI. There are many problems with this measure.
- (a) Open data file NY obesity and generate a scatterplot with years on the x -axis and percentage of people with BMI > 30 on the y -axis.
 - (b) Write a sentence explaining the meaning of the slope of your linear regression line in the context of the data.
 - (c) Use the model to predict obesity rates in 2020. Compare with your results in problem 2.
 - (d) Look at the residual plot of your least squares regression line. Are you confident that the data is linear? Explain.