



WEEK #2: DESCRIBING OUR DATA USING GRAPHS AND NUMBERS

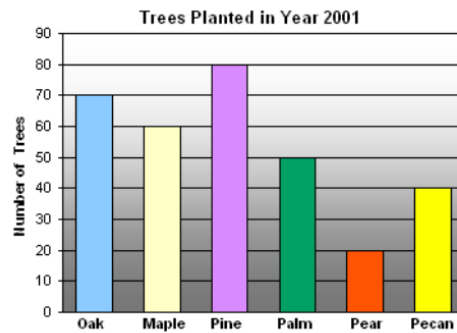
Problem 1. The Literary Digest magazine conducted a poll to predict the result of the 1936 Presidential election between Franklin Roosevelt (Democrat and incumbent) and Alf Landon (Republican). At the time, the poll was famous, because they had correctly predicted three successive elections. In 1936 they mailed questionnaires to 10 million people and asked how they planned to vote. The sampling frame was constructed from telephone directories, country club memberships, and automobile registrations. Only 2.3 million of those contacted returned their questionnaire. Based on their responses, the Literary Digest predicted that Landon would win, getting 57% of the vote. Instead, Landon got only 36% of the vote and Roosevelt won in a landslide. What happened?



Problem 2. Gastric freezing is a treatment for ulcers in the upper intestine. In a study published many years ago, patients with ulcers swallowed a deflated balloon with tubes attached, and then a refrigerated liquid was pumped through the balloon for an hour. The idea is that cooling the stomach will reduce its production of acid and relieve ulcers. Results from this study showed that about $1/3$ of the subjects improved. What is wrong with this study? How could this experiment be improved?



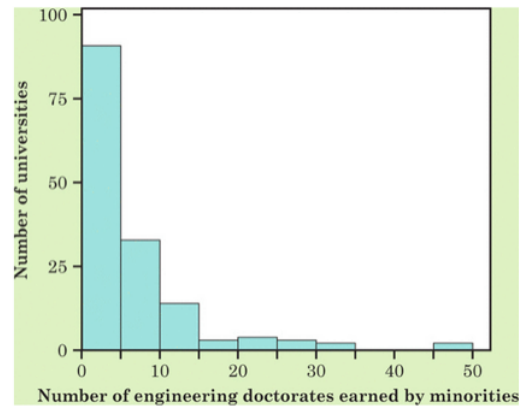
Problem 3. The following graph shows the different types of trees planted in Spring Hill City in 2001.



- What type of graph is this?
- What is the mode?
- Approximately how many Pear trees were planted?
- Approximately how many total trees were planted?



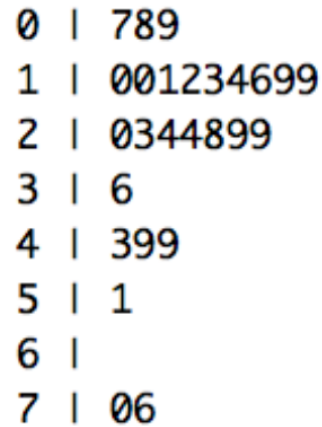
Problem 4. The figure below is a histogram of the number of minority students (black, Hispanic, Native American) who earned doctorate degrees in engineering from each of 152 universities in the years 2000 through 2002.



- About how many universities had between 0 and 10 engineering doctorates earned by minorities?
- How would you describe the shape of this histogram?
- Based only on this histogram, does it appear as if there are any possible outliers?



Problem 5. Alec is teaching an introductory math class and just gave an exam. The stemplot below shows the test scores for the first exam.

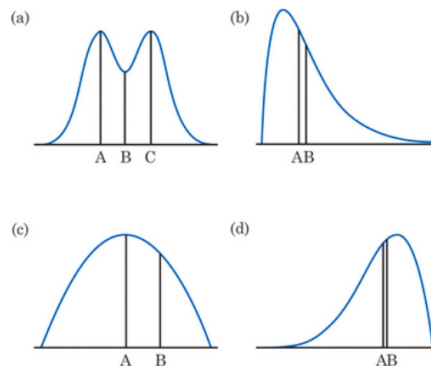


- What is the minimum?
- What is the first quartile (Q1)?
- What is the median?
- What is the third quartile (Q3)?
- What is the maximum?
- What is the interquartile range (IQR)?



Problem 6. Kyla has a coin collection that she started with her grandfather. She specifically collects quarters. Most of her quarters are fairly new (made between 2000 and 2017) but she has quite a few quarters from between 1950 and 2000. She even has a few quarters made as early as 1927. If you were to describe the distribution of the years in which Kyla's quarters were made, would you expect this distribution to be left-skewed, right-skewed, or symmetric?

Problem 7. The following diagram shows for different density curves, each with 2-3 points marked. For each curve, identify which points represent the mean and the median (some points may not be used).





Problem 8. Lauren's professor gives 15 point quizzes most weeks, for a total of 12 quizzes over the course of the semester. Lauren's scores are: 15, 11, 12, 13, 9, 14, 14, 12, 10, 9, 15, and 2. Are any of her scores potentially outliers?

Problem 9. Anna Grace is 2 years old and is 33 inches tall. Her doctor tells her parents that Anna Grace is in the 24.1st percentile for height. What does this mean?



Problem 10. Suppose that John's course has three midterms. His scores are: 82, 97, and 85.

- a. What is the mean?
- b. What is the variance?
- c. What is the standard deviation?
- d. What would John's midterm scores need to look like in order for the standard deviation to be zero?



Problem 11. For the following data sets, describe how the mean, median, IQR and standard deviation would change between the two data sets.

- a. Data Set 1: 3, 5, 6, 7, 8; Data Set 2: 3, 5, 6, 7, 20
- b. Data Set 1: 3, 3, 3, 3, 3, 3, 3; Data Set 2: 3, 3, 3, 3, 3, 3, 15

Problem 12. The midrange of a distribution is defined as the average of the minimum and the maximum of that distribution. Is this statistic robust to outliers and extreme skew? Explain your reasoning.