Welcome to AP Statistics! Before we officially kick off in August, I would like for you to get the ball rolling with the items described below. The purpose of this assignment is for you to review some basic graphs that you probably studied before. AP Statistics is a rigorous college-level course that demands a rigorous pacing, so by getting things started now will allow us to go more in-depth right away in August. The required portions do not require you to purchase anything and should require about 3 to 4 hours of your time. The optional book recommended should only cost about $10.

**REQUIRED**

1) Go to: [https://www.learner.org/courses/againstallodds/unitpages/index.html](https://www.learner.org/courses/againstallodds/unitpages/index.html)
   - Watch the first 5 “Against All Odds” video units (each are approximately 10 minutes long)
     - Video 1: What Is Statistics?
     - Video 2: Stemplots
     - Video 3: Histograms
     - Video 4: Measures of Center
     - Video 5: Boxplots
   - Do the exercises found on pages 2 – 10 of this document. For additional practice, explore and use the Interactive Tools and Student Guides available on this website. Your learning objectives are:
     a) Gain beginning understanding of why we should study statistics!
     b) Given a set of data, accurately create a well-labeled and appropriately scaled stemplot, histogram, and boxplot.
     c) Calculate and/or describe shape, center, and spread from a stemplot, histogram, or boxplot.

2) Go to: [https://apstudent.collegeboard.org/apcourse/ap-statistics](https://apstudent.collegeboard.org/apcourse/ap-statistics)
   - As you explore this site, please answer the questions on page 10 of this document.
   - Your learning objectives are:
     a) Determine whether statistics are important to the career areas and college majors that you are most interested in.
     b) Gain a broad overview of the AP Statistics course and the exam.

**OPTIONAL**

3) If you enjoy reading, I recommend “The Cartoon Guide to Statistics” by Larry Gonick and Woollcott Smith. I think you’ll enjoy reading this!

Please be prepared to turn in pages 2 through 10 of this document on the second day of school. This will be counted as a 50 point homework assignment for the fall semester. Also, there will be a 50 point quiz on this day for you to demonstrate what you’ve learned through this assignment.
After watching each video, answer the questions below.

**Video #1: What Is Statistics?**

1. What is statistics? Why is it important to study statistics?

**Video #2: Stemplots**

1. List some of the variables that were taken on soldiers for the sizing data bank.

2. What was the overall shape of the distribution of soldiers’ foot lengths? About where was the center of the distribution?

3. What variable was used to measure fuel economy on Toyota’s line of vehicles?

4. Focus on the stemplot of fuel economy for Toyota’s 2012 line. What new information became evident (or more clear) when the stem was expanded?
5. What was learned from back-to-back stemplots about the change in fuel economy in Toyota’s vehicle line from 1984 to 2012?

6. Below are the number of home runs that Babe Ruth hit in each of his 15 years with the New York Yankees, 1920 – 1934.

```
54  59  35  41  46  25  47  60  54  46
49  46  41  34  22
```

a. Make a stemplot of the home run data. Then use your stemplot to answer questions (b) and (c).

b. Describe the shape of the distribution. Is it roughly symmetric or not? Is it unimodal (single peak) or multimodal (more than one peak)?

c. What is the center (this is the number of home runs the Babe hit in a typical year)?

d. Ruth’s record of 60 home runs in 1927 stood for more than 30 years. Is 60 an observation that falls outside the pattern of the other observations and hence could be considered an outlier?
**Video #3: Histograms**

1. The video opens by describing a study of lightning strikes in Colorado. What variable does the first histogram display?

2. In this lightning histogram, what does the horizontal scale represent? What does the vertical scale represent?

3. Was the overall shape of this histogram symmetric, skewed (left or right?), or neither?

4. Why were a few values in the second lightning histogram called outliers?

5. When you choose the classes for a histogram, what property must the classes have if the histogram is to be correct?

6. What happens to a histogram if you use too many classes? What happens if you use too few?
7. The duration of 40 phone calls (in minutes) for technical support is given below.

<table>
<thead>
<tr>
<th>Duration (minutes)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 6</td>
<td></td>
</tr>
<tr>
<td>6 – 12</td>
<td></td>
</tr>
<tr>
<td>12 – 18</td>
<td></td>
</tr>
<tr>
<td>18 – 24</td>
<td></td>
</tr>
<tr>
<td>24 – 30</td>
<td></td>
</tr>
<tr>
<td>30 – 36</td>
<td></td>
</tr>
<tr>
<td>36 – 42</td>
<td></td>
</tr>
<tr>
<td>42 – 48</td>
<td></td>
</tr>
<tr>
<td>48 – 54</td>
<td></td>
</tr>
</tbody>
</table>

a. Complete the **frequency distribution table** for the call duration data.

b. What percentage of phone calls lasted less than 12 minutes?

c. What percentage of calls lasted a half hour or more?

d. Represent the frequency distribution with a well-labeled and appropriately scaled **histogram**. Use a percent scale on the vertical axis.

e. Describe the **shape** of the distribution. Are there any gaps in the data? Outliers?
**Video #4: Measures of Center**

1. What variable is examined in comparing men and women workers at the beginning of the video?

2. Would you describe the shape of the distribution of men’s weekly wages as symmetric, skewed to the left or skewed to the right?

3. What is the most important difference between the distributions of weekly wages for men and for women?

4. Would a few very large incomes pull the mean of a group of incomes up, down, or leave the mean unaffected?

5. Would a few very large incomes pull the median of a group of incomes up, down, or leave the median unaffected?
6. A student often orders french fries at a local fast-food place. She keeps track of the number of french fries in each small bag she buys. Here are her counts:

<table>
<thead>
<tr>
<th>42</th>
<th>47</th>
<th>49</th>
<th>58</th>
<th>43</th>
<th>47</th>
<th>44</th>
<th>38</th>
<th>38</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>40</td>
<td>46</td>
<td>54</td>
<td>45</td>
<td>45</td>
<td>51</td>
<td>35</td>
<td>46</td>
<td>37</td>
</tr>
<tr>
<td>46</td>
<td>40</td>
<td>43</td>
<td>49</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**a.** Calculate the **mean** and **median** for these data. Show how you computed these values.

**b.** Make a well-labeled **stemplot** for the data above. Describe the overall shape of the distribution. Are there any outliers?

**c.** Do you prefer the mean or the median as a brief description of the center of this distribution? Why?
Video #5: Boxplots

1. What variable is used to compare different brands of hot dogs?

2. What name do we give to the value for which one-quarter of the data values falls at or below it?

3. What numbers make up a five-number summary?

4. How do you calculate the interquartile range?

5. Boxplots show that poultry hot dogs as a group differ from all-beef hot dogs. Compare the distribution of calories between the two types of hot dogs. That is, discuss the similarities and differences about shape, center, and spread.

6. A consumer testing laboratory measured the calories per hot dog in 20 brands of beef hot dogs. Here are the results:

   186  181  176  149  184  190  158  139  175  148  
   152  111  141  153  190  157  131  149  135  132

   a. Find the five-number summary of this distribution. Explain how you arrived at your answer.
b. Compute the range and interquartile range. Explain what these numbers tell you about the variability in calories in different brands of all-beef hot dogs.

c. Would a beef hot dog with 175 calories be in the top quarter of the data? Support your answer.

d. Draw a basic boxplot for the calories per hot dog.

e. In which quarter – the first, second, third, or fourth – are the data most concentrated? Explain how you can answer this question based on the boxplot from (d).

f. In which quarter – the first, second, third, or fourth – is the data most spread out? Explain how you can answer this question based on the boxplot from (d).

g. If a data value is more than $1.5 \times \text{IQR}$ below the first quartile or more than $1.5 \times \text{IQR}$ above the third quartile, it is considered an outlier. Should any of the calorie counts for the beef hot dogs be classified as outliers? Explain.
**h.** Make a stemplot of the calories in the sample of beef hot dogs. What do you learn from the stemplot that you could not learn from the boxplot?

---

**After exploring the website about the AP Statistics course and exam, answer the questions below.**

1) “Course Overview”: How many career areas and college majors are listed as a possible future following AP Statistics? Skim through both lists. Which ones are you most interested in?

2) Click on “View Course Details” tab at top. What are the four broad conceptual themes of AP Statistics and about what percent of the course/exam is based on each of them?

3) Click on “About The Exam” tab at top. Briefly describe the two parts of the AP Statistics exam. How many multiple choice questions are there, and how much time do you get? How about for the free-response questions? Free-response question #6 is known as the “Investigative Task”. What about it sets it apart from the other questions?