Homework №1

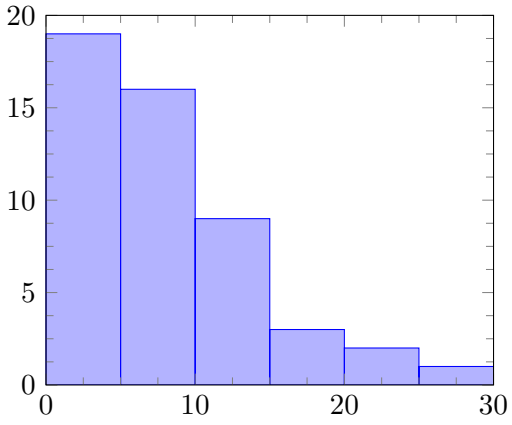
Author: David Oniani
Instructor: Dr. Eric Westlund

February 9, 2019

1.6 Since the max percentage is 27.1, we will have 6 bars. Notice that we have:

19 states in the category from 0% to $< 5.0\%$
16 states in the category from 5.0% to $< 10.0\%$
9 states in the category from 10.0% to $< 15.0\%$
3 states in the category from 15.0% to $< 20.0\%$
2 states in the category from 20.0% to $< 25.0\%$
1 state in the category from 25.0% to $< 30.0\%$

Also, notice that $19 + 16 + 9 + 3 + 2 + 1 = 50$ (so hope I did not miscategorize any state). Then the histogram for this data will be:



1.8 It is easy to see that the distribution is skewed and not symmetric. Particularly, this is the right-skewed distribution since its tail is to the right. The midpoint seems to be 15. The percent of state population born outside the united states ranges from 1.3% to 27.1%, which shows considerable variability. There are no states with unusually large or small percent of foreign-born residents since there are no gaps in the histogram.

1.10 (a) The stemplot will look like this:

5	9																								
6	2	3	7	8	8																				
7	1	1	2	4	4	4	5	6	6	6	6	7	7	7	8	8	8								
8	0	0	0	1	1	2	2	3	3	3	3	3	3	3	4	4	6	6	6	6	6	6	7	7	8

From this stemplot, it is easy to see that the midpoint is 80 and that the variability is from 59% to 88%.

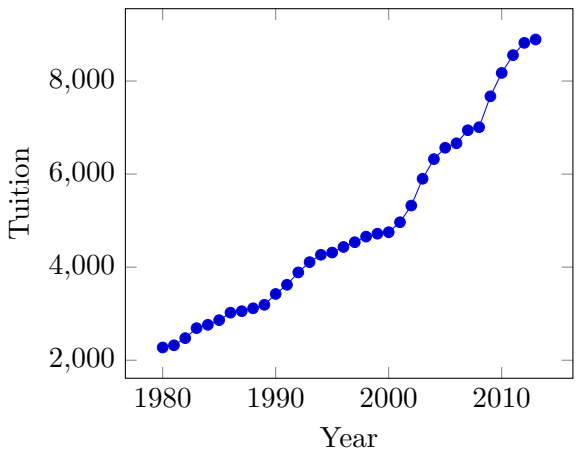
(b) The stemplot will look like this:

5															
5	9														
6	2	3													
6	7	8	8												
7	1	1	2	4	4	4									
7	5	6	6	6	6	7	7	7	8	8	8				
8	0	0	0	1	1	2	2	3	3	3	3	3	3	3	4	4
8	6	6	6	6	6	6	7	7	8							

In this stemplot, the shape of the distribution is a lot clearer (compared to histogram in Figure 1.5) to see. Though, it actually does give us a histogram if we put bars on top of these numbers.

1.12

(a) The time plot will look like this:

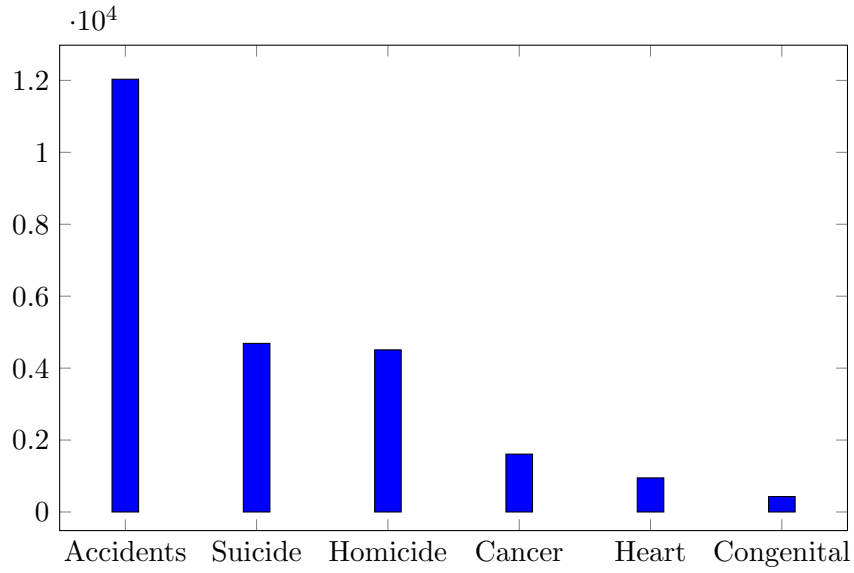


(b) The overall pattern is that over the years, the tuition increases.

(c) From 2001 to 2004 and from 2009 to 2012, we have periods of rapid increase. No other deviations are present.

(d) The percent increase would be better. Because it would also show us the how the change over time itself changes.

1.27 (a) The bar chart will look like this:



Notice that $0, 0.2, 0.4, 0.6, 0.8, 1, 1.2$ are all times 10^4.

(b) We would need to know how many people died from other causes since one cannot construct a pie chart without knowing what is the "whole picture."

1.28 They come from Mexico, Puerto Rico, Cuba, Central America, South America, and other regions. If my eyeballs tell the truth, then Mexicans must be somewhere around 60% ($\frac{3}{5}$ of the pie) of Hispanics. And approximately 10% ($\frac{1}{10}$ of the pie) of Hispanics are Puerto Rican.

1.33 2. Are you right-handed or left-handed? – This would go with graph (b) since, generally speaking, there are more right-handed people than left-handed ones.

1. Are you female or male? – This would go with graph (c). This is due to elimination. If it is not graph (b), it must be graph (c) since being a male or female is a binary option and all other graphs have more than 2 bars.
3. What is your height in inches? – The graph must look sort of symmetric since, generally speaking, there always is a great diversity in height (some are tall, some are average, some are short). The only symmetrically-looking graph is (d). Therefore, graph (d) seems to be the best option here.

4. How many minutes do you study on a typical weeknight? – Due to the elimination, we are left with the graph (a) which indeed makes sense since not a lot of people will study for a very long time resulting in the right-skewed distribution.

1.36 (a) A country with the bigger population would have a bigger CO_2 emission rate.

(b) I will round it to the nearest tenth (although rounding it to the nearest hundredth/t-housandth/etc. is more precise, it is super inconvenient and a lot of work too). This is how the stemplot looks like:

0	1	1	2	2	3	3	4	5	5	9	9
1	6	6	7	7	8						
2	2	6									
3	3	7	8								
4	4										
5	6	9									
6	2	6	7								
7	7	9									
8	3										
9	1	2	2								
10											
11	5										
12	2										
13											
14	6										
15											
16											
17	6										

where $0 \mid 1 = 0.1$, $2 \mid 6 = 2.6$ etc.

It is easy to see that the distribution is right-skewed. The center will be at 2 (there are 37 datapoints in total and the 18^{th} is 26). Variability is from 0.1 to 17.6. The US (17.6) and Canada (14.6) are the outliers.

1.38 (a) They mean that there were no improvements at all and that the task completion time worsened.

(b) The back-to-back stemplot looks like this:

Treatment		Control
	-0	8
1 2 2	-0	3 1
	0	5 5 6 7 7 8 9 9 9 9
8 8 8 7 7 6	0	2 3 3 4 4
4 3 3 3 2	1	5
8	1	1
4 4 2 1	2	3
8	2	3
3	3	

Note that numbers like 85 are rounded to 90. In other words, anything that ends with 0, 1, 2, 3, 4 gets rounded DOWN to 0 and anything that ends with 5, 6, 7, 8, 9 is rounded UP to 0.

(c) The midpoint for treatment is 130 and the midpoint for control is 90. No, it does not seem so. This is since $130 > 90$. Hence, control group did better. This is still somewhat inaccurate estimate since the problem gives us a median without saying whether the previous distribution was skewed or symmetric.

1.43 (a) Graph (a). In other words, if we drew an approximation lines for both graphs, then the slope of the line for graph (a) would be bigger than that for graph (b).

(b) The increase in tuition from 2000 to 2013 seems to be the same for both graphs. It is approximately $\$6,000$. I think that the graphs are for the same dataset, it is just the scaling on the vertical axis that is different.