Stat1600 Solution to Midterm #1, Form A

1. Fred is in a weekly golf league. His scores from the last $n = 21$ weeks are shown in the stem and leaf plot below:

```
-------------
Leaf Unit = 1

 7  6778
 8  012349
 9  01255
10  0112
11  01
-------------
(So, for example, the first observation under stem 11 is read as 110
 and the first observation under stem 7 is read as 76)
```

(a) (**5 points**) In addition to the Stem-and-Leaf Plot, what other type(s) of data presentation would be most appropriate for this data? Explain your answer. Your choices are: (Bar Chart, Pie chart, Histogram, Dotplot, Box-and-Whisker Plot)

Answer: Histogram, Dotplot, and Box-and-Whisker Plot can also be used for such (numerical) data.

(b) (**5 points**) After examining the Stem-and-Leaf Plot, what seems to be the shape of this data?

Answer: Right skewed

Reason: Long right tail

(c) (**10 points**) Give the five-number summary then construct a boxplot (with whiskers extending to the two extremes, MIN and MAX). (**Hint**: How do you construct a sorted list of the data from the stem-and-leaf plot above?)

MIN and MAX: MIN=76, MAX=111

Median: (show your work)

$.5(n+1) = .5 \times 22 = 11$ and hence

$$MED = (\text{11th ordered value}) = 90.$$

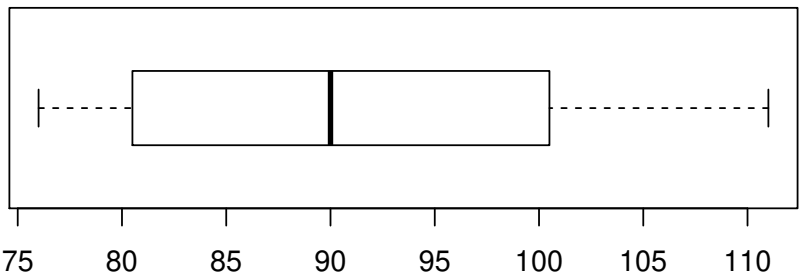
Quartiles, Q_1 and Q_3: (show your work)

$.25(n+1) = .25 \times 22 = 5.5$ and hence

$$Q_1 = \frac{(\text{5th ordered value}) + (\text{6th ordered value})}{2} = \frac{80+81}{2} = 80.5$$

$$Q_3 = \frac{(\text{5th largest value}) + (\text{6th largest value})}{2} = \frac{101+100}{2} = 100.5$$

boxplot:



2. The high temperature on Jan 30th (in °F.) for the last 5 years in a northern U.S. city is as follows:

```
25,31,23,24,27
```

(a) (**8 points**) Calculate the mean, $\overline{X}$ of this data.
ANSWER: 26

CALCULATION/REASON: (show your work)

$$\overline{X} = \frac{25+31+23+24+27}{5} = 26$$

(b) (**8 points**) Calculate the median of this data.

ANSWER: 25

CALCULATION/REASON: (show your work)

The sorted list of the data:

`23,24,25,27,31`

$.5(n+1) = .5 \times 6 = 3$ and hence

$$MED = (\text{3th ordered value}) = 25.$$

(c) (**10 points**) Let's compare the summary statistics of temperature for two cities during the past year.

Statistics	City A	City B
Mean	61°F	60°F
Standard Deviation	5°F	30°F

Based on the information provided in the table, would you expect one of these cities to have more extremes in their temperatures than the other?

ANSWER: Yes

REASON:

City B has more extremes than City A since the former has much larger standard deviation.

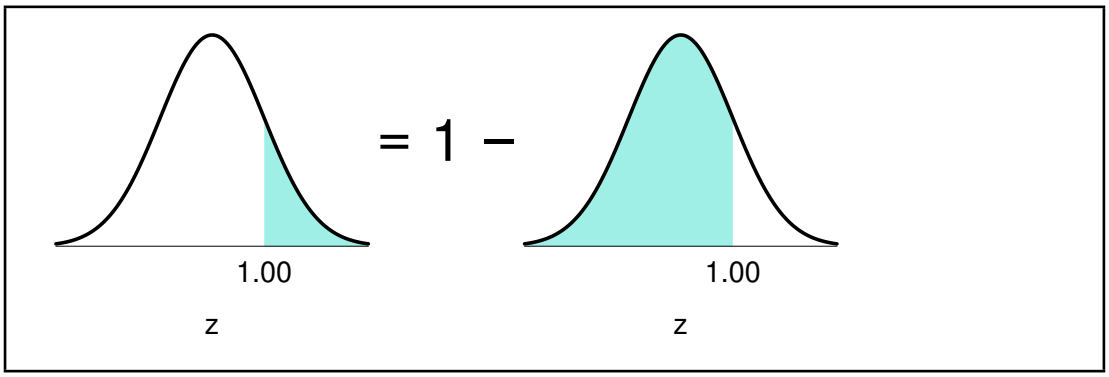
3. Let's assume that the height of males in the United States follows a Normal Distribution with a mean of 70 inches and a standard deviation of 3 inches. Given this information, use the Z-Table provided to answer the following questions:

(a) (**8 points**) What percentage of men are over 73 inches tall?

Note that the z-score of 73 is

$$z = \frac{73 - 70}{3} = 1.00$$

Hence, the chance that a man is over 73 inches tall can be computed as the area under the Z curve to the right of 1.00 which equals 1−0.8413 =0.1587 (see figures below). That is, 15.87%.



(b) (**8 points**) What percentage of men are between 64 and 76 inches?

Note that the z-score of 64 is

$$z = \frac{64 - 70}{3} = -2.00$$

and that the z-score of 76 is

$$z = \frac{76 - 70}{3} = 2.00.$$

Hence the percentage of men between 64 and 76 inches is the area (in percentage) under the Z curve bounded by -2.00 and 2.00 which is approximately 95% according to the empirical rules.

(c) (**8 points**) A male basketball player is 76 inches tall. What percentile is he in? (In other words, what percentage of males are shorter than he is?)

Note that the z-score of 76 is

$$z = \frac{76 - 70}{3} = 2.00$$

Hence the fraction of men that are shorter than he is can be computed as the area under the Z curve to the left of 2.00 which is 0.9772 from Z table. That is, The basketball play is at 97.72th percentile.

4. A recent survey, conducted by the Modern Language Association asks the following question: **Other than English, what other language(s) are spoken in your home?** A random sample of $n = 20$ households provided the following results:

Spanish Chinese Spanish Spanish Spanish
Chinese German Spanish Spanish French
Chinese Spanish Chinese Japanese Spanish
Spanish Chinese Chinese Chinese Spanish

(a) (**5 points**) What type of data is being analyzed here (Nominal, Ordinal, Interval, or Ratio? Explain your answer.

The data values are of nominal data type since there exists no sense of ordering.

(b) (**5 points**) What type(s) of data presentation would be most appropriate for this data? Explain your answer. Your choices are: (Bar Chart, Pie chart, Stem-and-Leaf Plot, Histogram, Dotplot, Box-and-Whisker Plot)

ANSWER:

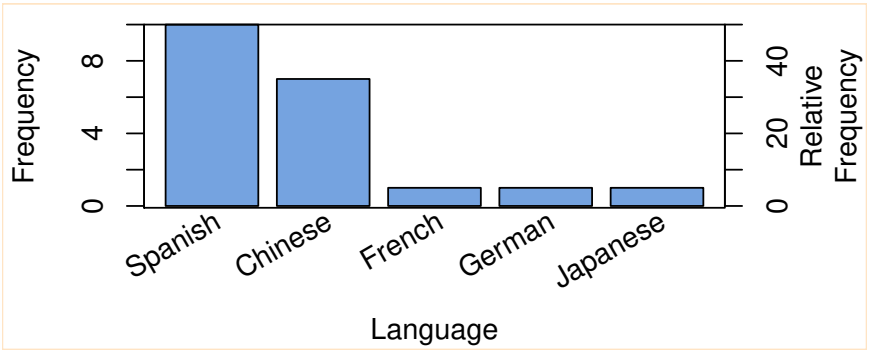
Choices that can be used are bar chart and pie chart. Others can not be used since the data set is nominal.

(c) (**10 points**) Construct a frequency distribution of this data.

	Frequency
Spanish	10
Chinese	7
French	1
German	1
Japanese	1
TOTAL	20

(d) (**10 points**) Construct a relative frequency table of this data.

	Frequency	Rel. Freq (%)
Spanish	10	50
Chinese	7	35
French	1	5
German	1	5
Japanese	1	5
TOTAL	20	100



5. (**5 points**) Extra Credit Problem
Assume we have a data set that includes the Telephone Area Codes of our customers. Some examples in the data set include 269, 616, 989 etc. Clearly these are numbers, but are considered nominal (categorical) data. Discuss or prove why these values are actually nominal (categorical) data and should not be analyzed as numeric data.

Although telephone area codes come as number, there exists no sense of ordering in the area codes. That is telephone area codes are nominal data values and should by no means be analyzed as numeric data. An average of, say 269, 616, 989, is 624.67. An area code of 624.67? That is located in the 'average' place? That is absurd!