Show **all** of your work on this assignment and answer each question fully in the given context. Remember to staple your assignment! Failure to do so will result in a five point deduction from your total score.

1. The 1997 season for NCAA football saw a split national title—the final AP poll gave the title to the U. of Michigan Wolverines while the final Coaches' Poll awarded their title to the U. of Nebraska Cornhuskers. This split championship was the tenth such occurrence in the previous 25 years, sparking a discussion about the effectiveness of the Bowl Alliance. In June 1998, the football conferences announced their solution, the Bowl Championship Series (BCS) system (notice: the BCS system created it's own share of controveries was replaced by the College Football Playoff). Below are the points scored per game during the 1997 season for the two schools (The first Nebraska game, 59-14 win vs. Akron was omitted. Source: *Wikipedia*).

Wolverines:	27	38	21	37	23	28	23	24	34	26	20	21
Cornhuskers:	38	27	56	49	29	35	69	45	77	27	54	42

   (a) Draw a back-to-back stem-and-leaf display of the two pssssint distributions. Put the Wolverines on the left side of the stem, and the Cornhuskers on the right side. Remember to give a key or legend (e.g., $1|4 = 14$), so the reader can interpret the display. (A back-to-back stem-and-leaf plot is illustrated in Figure 3.5 on page 70 of the course text)

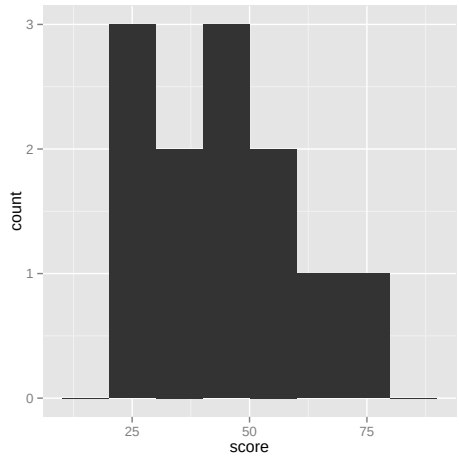
```
key: 2|7 = 27

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

   (b) Construct a frequency table for the **Cornhuskers'** point distribution. Your table should have to following column headings: points, frequency, relative frequency, and cumulative relative frequency. Choose a 10-point interval—large enough so the data are grouped together (there should be few intervals with zero or one data points), but enough intervals to adequately show features of the distribution (there should be more than three intervals).

Range	Frequency	Relative Frequency	Cumulative Relative Frequency
20-29	3	$3/12 = 0.250$	$3/12 = 0.250$
30-39	2	$2/12 = 0.167$	$5/12 = 0.417$
40-49	3	$8/12 = 0.167$	$8/12 = 0.667$
50-59	2	$10/12 = 0.167$	$10/12 = 0.833$
60-69	1	$11/12 = 0.167$	$11/12 = 0.917$
70-79	1	$12/12 = 0.167$	$12/12 = 1.000$

   (c) Draw a histogram for the **Cornhuskers'** point distribution using the frequency table.



(d) For **both** the Cornhuskers and the Wolverines, Calculate the first quartile, the median, and the third quartile using the quartile function $Q(p)$.

Nebraksa:

$$\begin{aligned} Q(.25) &= x_i + (n(p) - i + .5)(x_{i+1} - x_i) && \text{(since } np + 0.5 = 3.50\text{)} \\ &= x_3 + (12(.25) - 3 + .5)(x_{3+1} - x_3) && \text{(since } p = 0.25\text{)} \\ &= 29 + (3 - 3 + .5)(35 - 29) \\ &= 29 + (.5)6 \\ &= 32 \end{aligned}$$

$$\begin{aligned} Q(.5) &= x_i + (n(.5) - i + .5)x_{i+1} && \text{(since } np + 0.5 = 6.50\text{)} \\ &= x_6 + (12(.5) - i + .5)(x_{6+1} - x_6 && \text{(since } p = 0.5\text{)} \\ &= 42 + (6 - 6 + .5)(45 - 42) \\ &= 42 + (.5)3 \\ &= 43.5 \end{aligned}$$

$$\begin{aligned} Q(.75) &= x_i + (n(.75) - i + .5)x_{i+1} && \text{(since } np + 0.5 = 9.50\text{)} \\ &= x_9 + (12(.75) - 9 + .5)(x_{9+1} - x_9 && \text{(since } p = 0.75\text{)} \\ &= 54 + (9 - 9 + .5)(56 - 54) \\ &= 54 + (.5)2 \\ &= 55 \end{aligned}$$

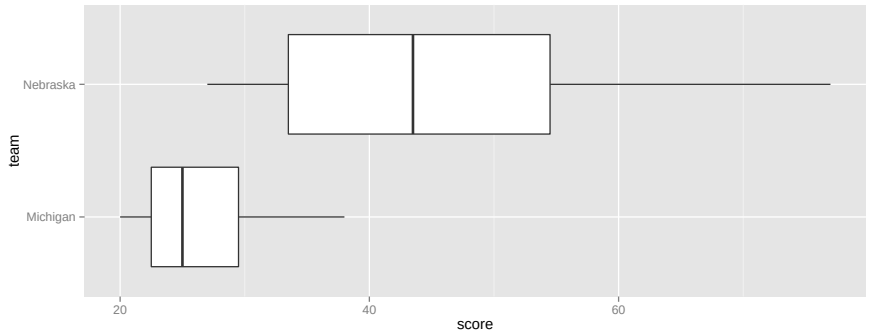
Michigan:

$$\begin{aligned} Q(.25) &= x_i + (n(p) - i + .5)(x_{i+1} - x_i) && \text{(since } np + 0.5 = 3.50\text{)} \\ &= x_3 + (12(.25) - 3 + .5)(x_{3+1} - x_3) && \text{(since } p = 0.25\text{)} \\ &= 21 + (3 - 3 + .5)(23 - 21) \\ &= 21 + (.5)2 \\ &= 22 \end{aligned}$$

$$\begin{aligned} Q(.5) &= x_i + (n(.5) - i + .5)x_{i+1} && \text{(since } np + 0.5 = 6.50\text{)} \\ &= x_6 + (12(.5) - i + .5)(x_{6+1} - x_6 && \text{(since } p = 0.5\text{)} \\ &= 24 + (6 - 6 + .5)(26 - 24) \\ &= 24 + (.5)2 \\ &= 25 \end{aligned}$$

$$\begin{aligned} Q(.75) &= x_i + (n(.75) - i + .5)x_{i+1} && \text{(since } np + 0.5 = 9.50\text{)} \\ &= x_9 + (12(.75) - 9 + .5)(x_{9+1} - x_9 && \text{(since } p = 0.75\text{)} \\ &= 28 + (9 - 9 + .5)(34 - 28) \\ &= 28 + (.5)6 \\ &= 31 \end{aligned}$$

(e) Construct side-by-side boxplots comparing the data for the Wolverines and the Cornhuskers. Make the axis range from 20 to 80 points with tick marks every 10 units. Based on the boxplots, are there *unusual* observations for either school?

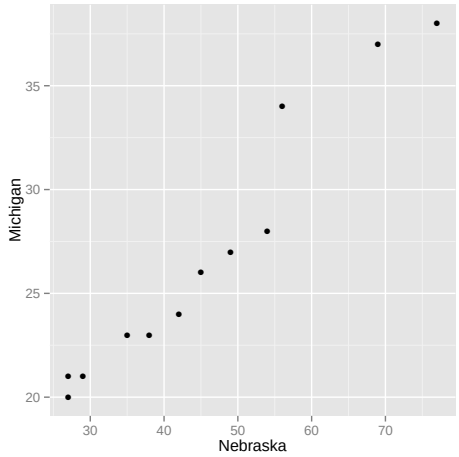


Based on these plots, there are no unusual observations for either of the two schools.

(f) Construct a quantile-quantile plot to compare the shape of the two distributions. Would you say that the plot appears to be linear? What does that indicate about the shapes of the two distributions?

Since both teams have the same number of games, a Q-Q plot can be constructed by simply plotting the pairs of the ordered scores.

	1	2	3	4	5	6	7	8	9	10	11	12
Nebraska	27	27	28	35	38	42	45	49	54	56	69	77
Michigan	20	21	21	23	23	24	26	27	28	34	37	38

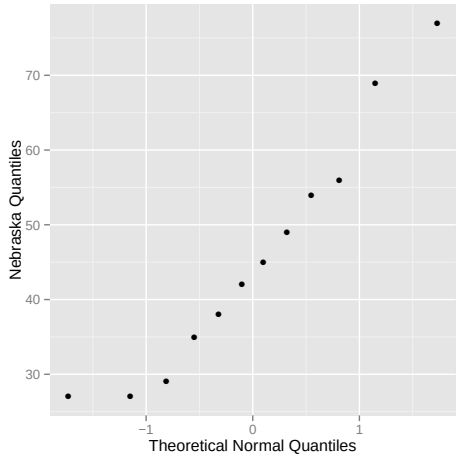


The plot does not look very linear. This would seem to suggest that the scores of the two teams do not share a common distribution (by which I mean they are spread out, or distributed, differently).

(g) Construct a theoretical quantile-quantile plot using a normal distribution as the theoretical normal probability plot for the Wolverines.The table below displays the quantiles of the theoretical normal distribution when $n = 12$:

p	1/24	3/24	5/24	7/24	9/24	11/24
$Q(p)$	-1.73	-1.15	-0.81	-0.55	-0.32	-0.10
p	13/24	15/24	17/24	19/24	21/24	23/24
$Q(p)$	0.10	0.32	0.55	0.81	1.15	1.73

Would you say that the plot appears to be linear? Does this imply the Wolverines football scores could have come from a normal distribution?



Here is the plot:

This plot does not appear to be linear - the lower values of the theoretical distribution would need smaller scores for the points to truly resemble a line. Notice that this implies

Nebraska's low scoring games were still higher bell-shape would have created.

(h) Calculate the sample mean, sample variance, and sample standard deviation for each school. Clearly label your answers.

Hint: for the Wolverines $\sum_{i=1}^{12} x_i = 322$, $\sum_{i=1}^{12} x_i^2 = 9074$,

and for the Cornhuskers $\sum_{i=1}^{12} x_i = 548$, $\sum_{i=1}^{12} x_i^2 = 27900$.

For the Wolverines:

Mean: 26.8333333

Variance: 39.4242424

Sample standard deviation: 6.2788727

For the Cornhuskers:

Mean: 45.6666667

Variance: 261.3333333

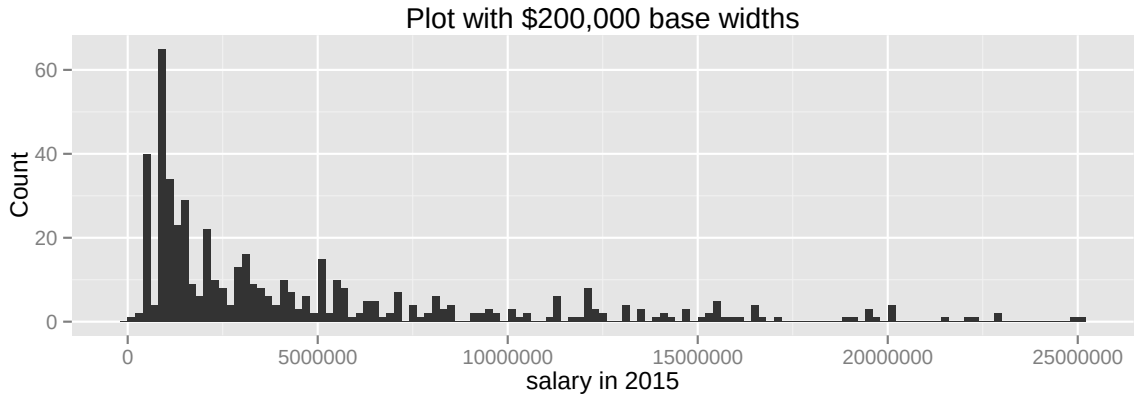
Sample standard deviation: 16.1658075

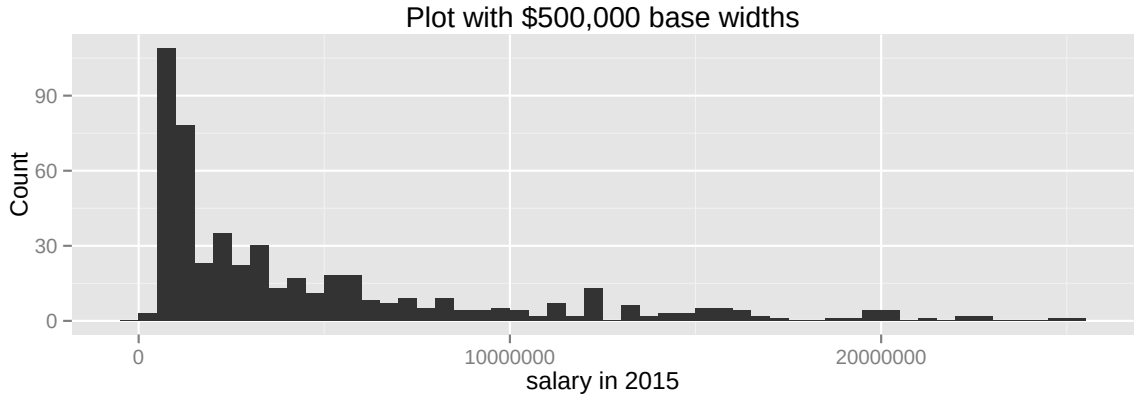
2. The NBA uses a salary cap in the hopes to keep the league competitive - the amount each team can spend on player salaries is limited so that no team can acquire all the best players. While this sounds simple in concept, the execution is often complicated and statistical evaluation of players' pay is a popular topic in sports journalism.

   The ten highest paying contracts in the 2015-2016 season are:

	player	team	salary2015
1	Kobe Bryant	LAL	25000000
2	Joe Johnson	BRK	24894863
3	LeBron James	CLE	22971000
4	Carmelo Anthony	NYK	22875000
5	Dwight Howard	HOU	22359364
6	Chris Bosh	MIA	22192730
7	Chris Paul	LAC	21468696
8	Kevin Durant	OKC	20158622
9	Derrick Rose	CHI	20093063
10	Brook Lopez	BRK	20000000

   Below are three histogram's summarizing the salary of each player.





   Note that the width of the intervals is the only important feature think that changes between

the two plots

(a) Which of the two histograms do you think does a better job of summarizing the data (meaning which plot helps you get a better picture of what is happening with salaries in the NBA)?
The second plot seems to be better at describing the overall shape of NBA salaries.

(b) Using the terminalogy from class and discussed in section 3.1.2, describe the distribution of the NBA salaries. Be sure to comment on the number of modes and symmetry.
There is very clearly a mode near the lower end of salaries. In fact the data is almost right skewed. An argument could be made that the data is unimodal, with a second, though much less prominent mode near $12,000,000.

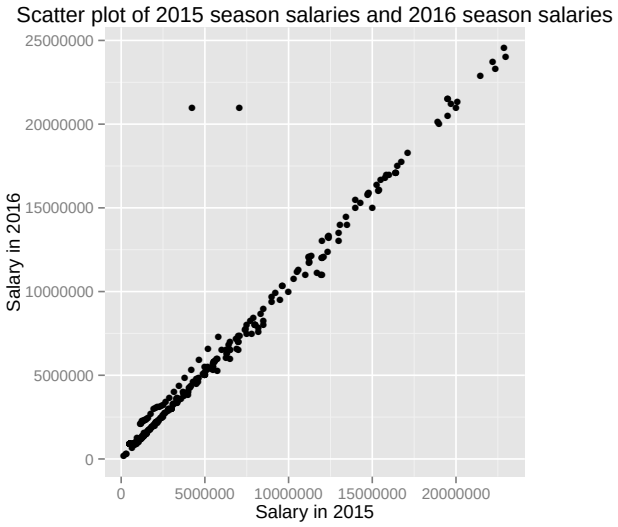
(c) Does the sample mean or sample median provide a more appropriate measure of center (location) of the distribution of 2015 NBA salaries? Explain briefly.
Because of the skew in the data, the median would likely provide a better measure of "where the middle" of the data is.

(d) Does the sample standard deviation or sample IQR provide a more appropriate measure of spread (variability) of the distribution of 2015 NBA salaries? Explain briefly.
IQR would be better than median in this case, since IQR does not change with large values.

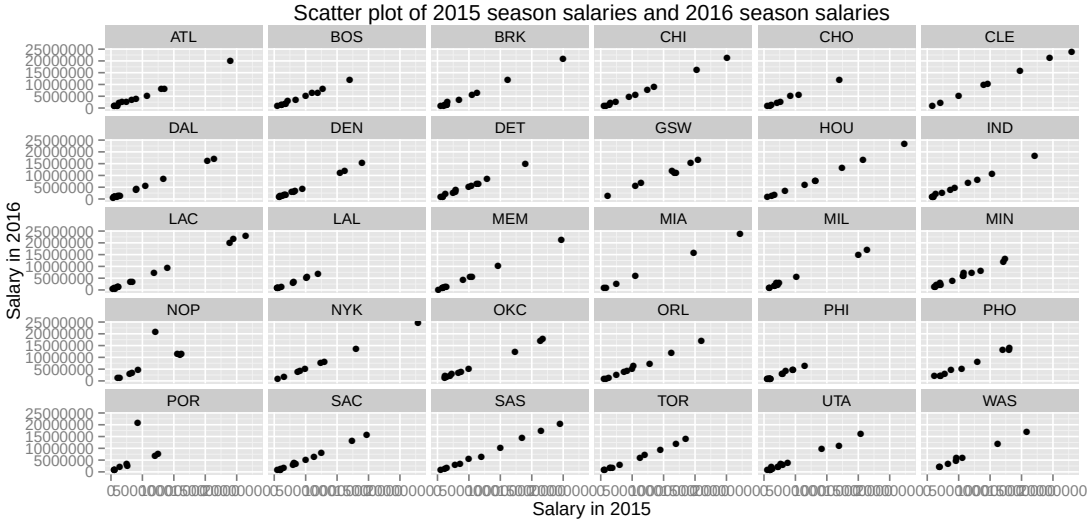
There is also information on the anticipated salary for players next season who are under contract at that point. Here are the top ten salaries for next year's players:

	player	team	salary2016
1	Carmelo Anthony	NYK	24559380
2	LeBron James	CLE	24004000
3	Chris Bosh	MIA	23741060
4	Dwight Howard	HOU	23282457
5	Chris Paul	LAC	22868828
6	Kevin Love	CLE	21500000
7	DeAndre Jordan	LAC	21500000
8	Derrick Rose	CHI	21323250
9	Marc Gasol	MEM	21200000
10	Damian Lillard	POR	21000000

We can create a scatter plot of the current salary and next seasons salary:



We can also look at each team's salary information:



(e) Describe the relationship between the 2015 season salary and the 2016 season salary.

Salaries the salaries seem to be very closely related (almost linearly). Though there are exceptions, it appears that what an NBA player earns in 2016 will be very much like what he earned in 2015.

(f) What sort of structural features/curiosities are present in this data set? Explain.

Here are a few:

- For most teams, the salary payed to a player in 2016 will be very much like the salary payed in 2015 (all the points fall near a line).
- There are two strong exceptions - one player for NOP (New Orleans) and one player for POR (Portland) are set to have huge increases in pay next season.
- Most teams seem to follow the "superstar" trend, where most of the players clump up near the lower end of salary, but a few players are payed a lot.

- Some teams - LAL, POR, PHI for example - have no one being paid highly in both 2015 and 2016. This could be because the player has no contracted 2016 salary (and thus no point on any graph). This would also explain why Kobe Bryant (the highest paid player in 2015) is not on the table for 2016.
- Some teams (curiously, very good teams) have more players in the range between superstar salaray and base level salary. Houston, Cleveland, and San Antonio seem to have a more "spread out" set of salaries.

3. **JMP Assignment.**

   Without laboring the point, computing is one of the most important parts of modern data analysis. A large part of data science simply wouldn't exist without the tools developed by scientists working at the intersections of computer science, mathematics, and statistics. Because of that, there will inevitably be parts of this course where a statistical computing tools are needed. SAS and R are the two main languages used by statisticians, with Python, Julia, F#, C++ and others making important contributions as well. SAS has a software called JMP ("Jump") that makes doing statistical analyses simpler - it is more powerful than Excel or your calculator but requires little in the sense of coding making the learning curve much lower. We will be using it this semester. There are labs in Snedecor Hall with the software pre-installed, but it is free for students and I encourage you to download a copy for yourself using the link below.

   Download: http://www.stat.iastate.edu/resources-2/software-sasjmpr/statistical-software-jmp/

   Additionally, you may want to consider the following tutorials (they are very helpful):

   Tutorials: http://web.utk.edu/ cwiek/201Tutorials/

   The tutorials cover the following topics:

   - Histogram and Box Plot
   - Stem and Leaf Plot
   - Normal Probability Plot and Goodness of Fit Test
   - Calculating Summary Statistics of Quantitative Data
   - Getting JMP Graphics into Microsoft Word

   For this problem I am asking you to:

   (a) Download and install `JMP` or find a computer with it already installed.
   (b) Take a screen shot once you have it open.