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)

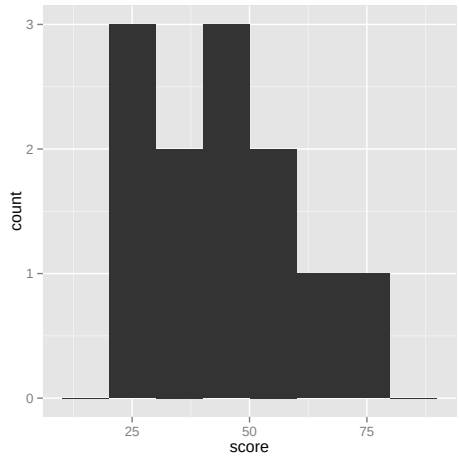
```
key: 2|7 = 27

                                | 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,

i. Calculate the first quartile, the median, and the third quartile using the quartile function $Q(p)$.
Nebraksa:

p	Nebraska Q(p)	Michigan Q(p)
.25	$Q(.25) = x_3$+ 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	$12/12 = 0.167$
$12/12 = 1.000$		

ii. 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?

```
qplot(score,team,data=teams,geom="boxplot",xlim=c(20,80))
## Error in eval(expr, envir, enclos):  could not find function "qplot"
```

iii. 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?

iv. 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?

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

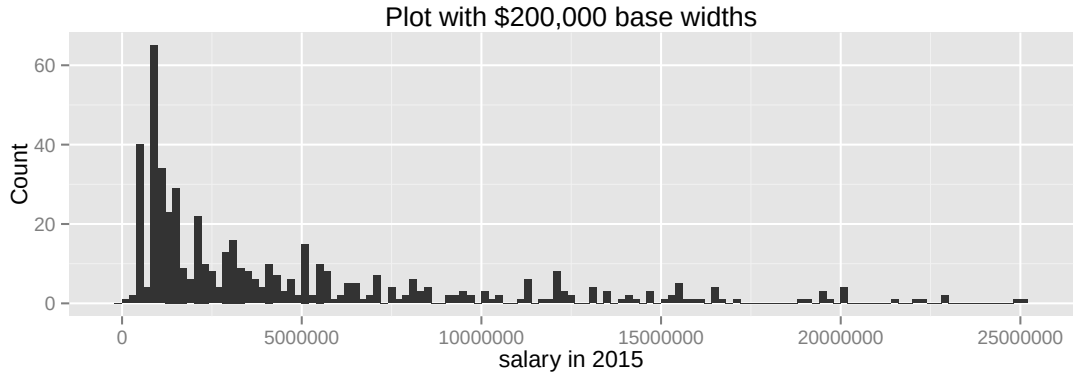
Hint: for the Wolverines $\sum_{i=1}^{12} x_i = 322$, $\sum_{i=1}^{12} x_i^2 = 9074$,

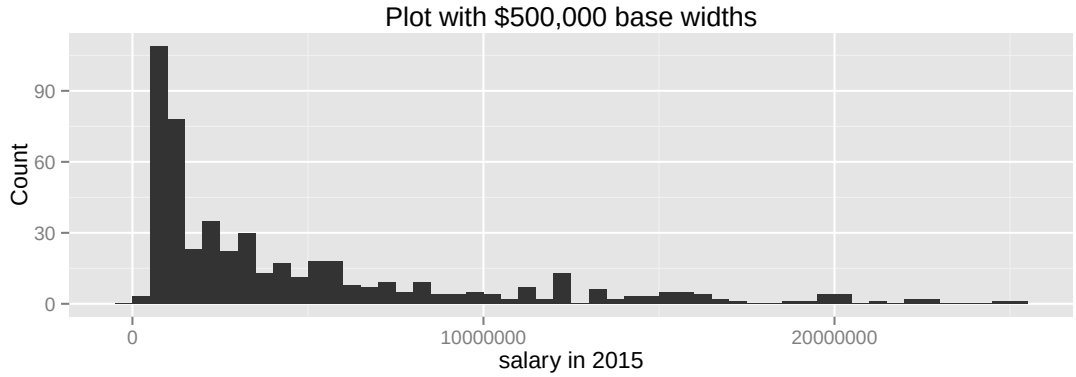
and for the Cornhuskers $\sum_{i=1}^{12} x_i = 548$, $\sum_{i=1}^{12} x_i^2 = 27900$.

(e) 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:

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





Note that the width of the intervals is the only important feature thing that changes between the two plots

i. Which of the two histograms do you thing 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)?

ii. Using the terminalogy from class and discussed in section 3.1.2, describe the distribution of yards per game. Be sure to comment on the number of modes and symmetry.

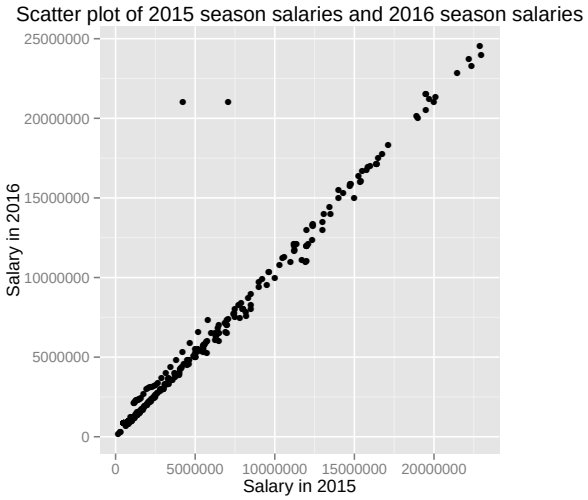
iii. Does the sample mean or sample median provide a more appropriate measure of center (location) of the distribution of 2015 NBA salaries? Explain briefly.

iv. 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.

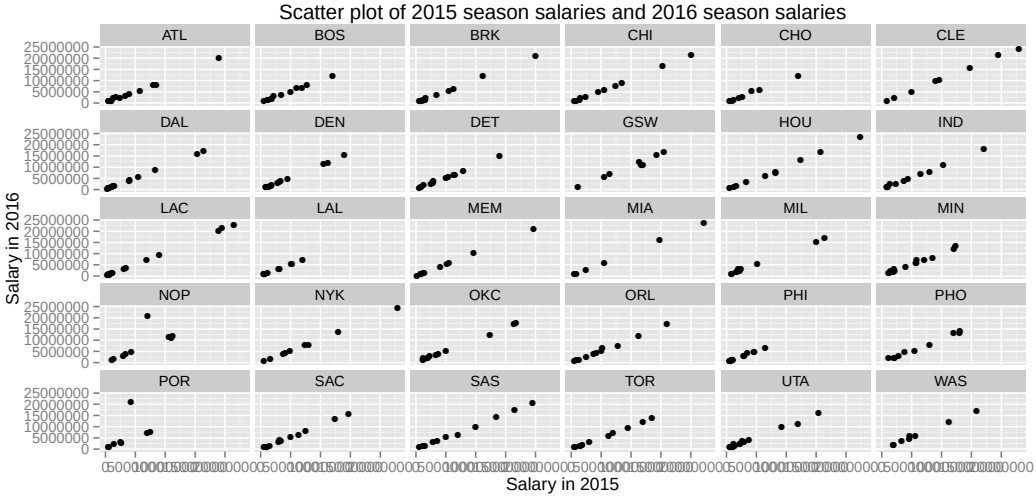
There is also information on the anticipated salary for players next season who are under contract at that point. Here are the top ten salaries for next year's players:

	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:



v. Describe the relationship between the 2015 season salary and the 2016 season salary.

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

(f) **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:

i. Download and install `JMP` or find a computer with it already installed.

ii. Take a screen shot once you have it open.