1. Create a Data Frame, 'CityData', using the two vectors that you've just created. Which city consists maximum temperature?
2. Add another variable to your data frame, 'population'. Which command did you use? Paste your command below.
3. Create a new data frame with the name of 'NewCityData' that contains two new cities and correcsponding variables with the same name as in previous data frame.
4. Combine both the data frames using rbind function and see paste the result.

Please download the "AnonymityPoll.csv" data file before starting this section. The file can be found in the Shared Google Drive folder.

1.1 ) How many people participated in the poll?

1.2 ) How many interviewees responded that they use a smartphone?

1.2.1 ) How many interviewees responded that they don't use a smartphone?

1.2.2 ) How many interviewees responded that they don't use a smartphone?

1.2.3 ) How many interviewees did not respond to the question, resulting in a missing value, or NA, in the summary() output?

1.3 ) Which of the following are states in the Midwest census region? (Select all that apply.)

Colorado

Kansas

Kentucky

Missourie

Ohio

Pensylvania

1.3.1 ) Which was the state in the South census region with the largest number of interviewees?

2.1 ) How many interviewees reported not having used the Internet and not having used a smartphone?

2.1.1 ) How many interviewees reported having used the Internet and having used a smartphone?

2.1.2 ) How many interviewees reported having used the Internet but not having used a smartphone?

2.1.3) How many interviewees reported having used a smartphone but not having used the Internet?

2.2) How many interviewees have a missing value for their Internet use?

2.2.1) How many interviewees have a missing value for their smartphone use?

2.3) Use the subset function to obtain a data frame called "limited", which is limited to interviewees who reported Internet use or who reported smartphone use. In lecture, we used the & symbol to use two criteria to make a subset of the data. To only take observations that have a certain value in one variable or the other, the | character can be used in place of the & symbol. This is also called a logical "or" operation.How many interviewees are in the new data frame?

Use the "limited" data frame that you have just created to answer all the subsequent questions.

3.1 ) Which variables have missing values in the limited data frame? (Select all that apply.)

Internet.Use

Smartphone

Sex

Age

State

Region

Conservativeness

Info.On.Internet

[Worry.About.Info](http://worry.about.info)

Privacy.Importance

Anonymity.Possible

Tried.Masking.Identity

Privacy.Laws.Effective

3.2 ) What is the average number of pieces of personal information on the Internet, according to the Info.On.Internet variable?

Your answer

3.3 ) How many interviewees reported a value of 0 for Info.On.Internet?

Your answer

3.3.1 ) How many interviewees reported the maximum value of 11 for Info.On.Internet?

Your answer

3.4 ) What proportion of interviewees who answered the[Worry.About.Info](http://worry.about.info) question worry about how much information is available about them on the Internet?

Your answer

3.5 ) What proportion of interviewees who answered the Anonymity.Possible question think it is possible to be completely anonymous on the Internet?

Your answer

3.6 ) What proportion of interviewees who answered the Tried.Masking.Identity question have tried masking their identity on the Internet?

Your answer

3.7 ) What proportion of interviewees who answered the Privacy.Laws.Effective question find United States privacy laws effective?

4.1 ) Build a histogram of the age of interviewees. What is the best represented age group in the population?

People aged about 20 years old

People aged about 40 years old

People aged about 60 years old

People aged about 80 years old

4.2 ) What is the largest number of interviewees that have exactly the same value in their Age variable AND the same value in their Info.On.Internet variable? In other words, what is the largest number of overlapping points in the plot plot(limited$Age, limited$Info.On.Internet)? (HINT: Use the table function to compare the number of observations with different values of Age and Info.On.Internet.)

4.3.1 ) What proportion of smartphone users who answered the Tried.Masking.Identity question have tried masking their identity when using the Internet?

Hint - Use tapply function to derive the results

4.3.2 ) What proportion of non-smartphone users who answered the Tried.Masking.Identity question have tried masking their identity when using the Internet?

Hint - Use tapply function to derive the results

4.3.3 ) Next week, we will begin to more formally characterize how an outcome variable like Info.On.Internet can be predicted with a variable like Age or Smartphone.