Name _____

HOMEWORK APRIL 6TH, 2017
DUE APRIL 7TH, 2017

Collect the height, weight, age, and gender of five different people. Try to get people of different ages! This information can come from your family, your friends, or celebrity information you find online! We will look at all of the different information you collected during our Biochemistry is Elementary sessions.

PERSON 1:
Height:_____ feet _____ inches
Weight: _____ pounds
Age: _____ years
Gender (circle): Male Female Rather Not Say

PERSON 2:
Height:_____ feet _____ inches
Weight: _____ pounds
Age: _____ years
Gender (circle): Male Female Rather Not Say

PERSON 3:
Height:_____ feet _____ inches
Weight: _____ pounds
Age: _____ years
Gender (circle): Male Female Rather Not Say

PERSON 4:
Height:_____ feet _____ inches
Weight: _____ pounds
Age: _____ years
Gender (circle): Male Female Rather Not Say

PERSON 5:
Height:_____ feet _____ inches
Weight: _____ pounds
Age: _____ years
Gender (circle): Male Female Rather Not Say

BIOCHEMISTRY IS ELEMENTARY

Your Name:_____

PROJECT SITE:

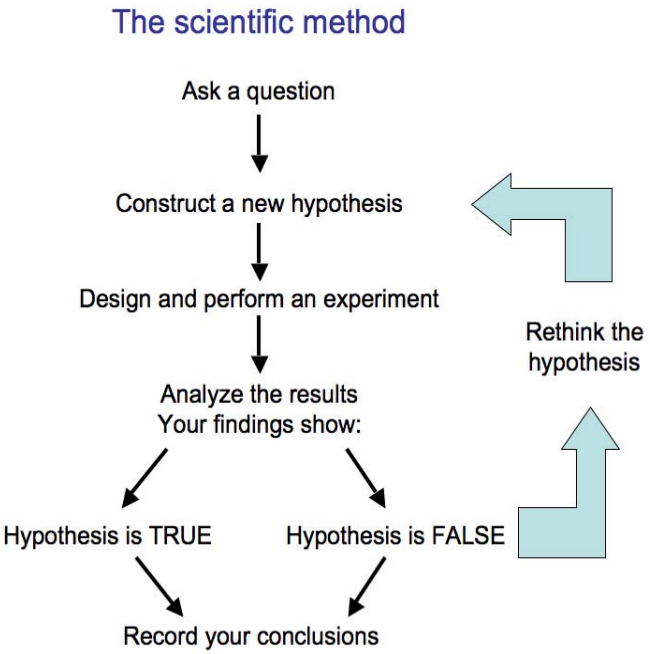
PROJECT SUPPORT: Tsao family gift to Colorado State University, & the National Science Foundation

Colorado State University

PROJECT LEADERS:
Drs. Cathy Radebaugh, Eric Ross & Laurie Stargell
Department of Biochemistry & Molecular Biology, College of Natural Sciences

SECTION 1 (CLASS 1)

PROJECT: What is Big Data? Collection and Analysis



Project: What is Big Data? Collection and Analysis
Concept: Introduction to big data, responsible data collection and analysis.
Created by: Hailey Conover Sedam, Nadia Sampaio, Zach Fox and Adam Heck

Introduction

Can you write down your height, your age and your gender? Sure you can. Can you write it down for everyone in your class? Or for everyone in your school? Or everyone in your town? Everyone in the country? Everyone in the world? And then write down their age, their favorite color, their favorite animal, their favorite TV programs, almost anything you can think of … for millions and millions of people?

That's too much, isn't it? You can't do it. No one can do it. But some computers can do this, and we call it **'big data'**. We can collect data about anything that we find interesting, anything that we are curious about and want to learn more. Data is all the stuff we are writing down and you can imagine, for millions, billions, gazillions of people it is really BIG.

What can we do with all this data? We can find patterns. For example, people making television programs may discover that children who like music really like drawing too, so they could make a new program about music and art. Or hospitals could discover that people who like different foods may get sick in different ways as they get older. That could help families and doctors look after our grandparents better.

(Based on Donald Farmer's definition of 'Big Data'. You can find his original text and the definitions proposed by other experts in: http://www.cbronline.com/news/big-data/analytics/13-ways-to-explain-big-data-to-a-five-year-old-4315589/2/)

Important Vocabulary:
Data:
Big Data: term that describes large and complex data sets.
Random sample: a selection that is chosen purely by chance, i.e. that does not favor any specific group.
Sampling: the act or process of selecting an appropriate sample.
Scatterplot: a graphic tool used to display the relationship between two things. Making a scatterplot will help us make predictions and to understand all of the valuable information from the data we collected.
Cluster: a group of similar data points.
Outlier: a data point that does not fit into a cluster or group.
Statistics: the study of data. How to collect, summarize and present it.
Correlation: a mutual relationship or connection between two things.

Today's Project:

Today we will analyze the height weight age and gender data that you collected and see if we can make something meaningful out of the large dataset your class made.

Last week you went home and collected data from five people. This week, we will combine the class' data to form a **big data** set. We will then analyze that big data set by graphing it on a **scatterplot** and then **clustering** data points into meaningful groups.

Statement of our question:

_____

_____

_____

Statement of our hypothesis:

_____

_____

_____

As scientists we need to record the material and equipment we will be using in this study. Please check off the following materials.

Materials and Equipment:

_____ 3 colored medium post-its per student with the data you collected

_____ 3 colored small post-its

Procedure for Big Data Experiment:

1. Last week you collected data from five people. Three of those people (data-points) were selected per student and were put into this clear box, and on this scatterplot.

Number of students in this class:

Description of your experimental group:

_____

_____

Record all of the information you can gather from the data in the clear box and the scatterplot.

_____

_____

Did the scatterplot help you uncover information that you couldn't see in the clear box? If so what did you discover?

_____

_____

2. Now, you will be given 3 post-its with the information you collected written on them. One sticky note per person sampled (data-points), pink for females and blue for males. The data given to you is identical to the data in the clear box and on the yellow scatter plot.
3. Graph your data: Place your 3 post-it notes over the yellow post-its in their appropriate location on the scatter plot. The y-axis is height and the x-axis is weight.

Make Pink or Blue circles on the scatter plot at the top of this page where you see **clusters** of pink or blue post-it notes.

Is there a correlation between gender and height and weight? If so, what is it?

_____

_____

Procedure for Age Clustering:

1. Take the small colored post-it notes and place them on your pink and blue post-it notes on the board based on the age of the individual.
   Green if the age is 50 years or older
   Purple if the age is 20 years- 50 years old
   Yellow if the age is 10 years- 20 years old
   Blue if the age is 0 years- 10 years old
2. Go to scatterplot on page 7 and circle any clusters of green, purple, yellow, or blue small post-it notes that you see.

Did clustering the data by age group on the scatterplot help you uncover information that you couldn't see before? If so what did you discover?

_____

_____

Using your Data to Make Prediction:

Write down the gender of your CSU teacher _____

Write down the age of your CSU teacher _____

Write down the height of your CSU teacher _____

Can you use the gender and age of you CSU teacher and the scatterplot you created today to predict your CSU teacher's weight?

Weight prediction _____

Actual weight _____

Summary Questions

1. Is our sample random? (Circle one)

   YES NO

2. Did height and weight correlate? (Circle one)

   YES NO

3. Did gender of the individual help cluster the data? (Circle one)

   YES NO

4. Did height and weight correlate with gender? (Circle one)

   YES NO

5. Did the age of the individual help cluster the data? (Circle one)

   YES NO

6. Did height and weight correlate with age? (Circle one)

   YES NO

7. Was your initial data predictive of the CSU Teacher's weight? (Circle one)

   YES NO

8. Were your hypotheses correct? (Circle one)

   YES NO

9. Did you enjoy performing this experiment?

   Not sure YES NO

10. Did you learn something about humans (like yourself) today? (Circle one)

    Not sure YES NO

Section 3 checklist:

When undertaking a study using the scientific method, it is important to make sure you performed all the necessary steps.

The following is a checklist to help you do this.

Checklist item	Answers should be "yes", if not go back and complete the steps.
1. Have you recorded the hypothesis you are testing?	
2. Have you worded the hypothesis so that it can be tested in your experiment?	
3. Did you make a list of the materials you used?	
4. Do you think another individual can repeat your experiment based on the experimental procedure?	
5. Did you collect your data using a table?	
6. Did you repeat the experiment?	
7. Were you consistent, careful and accurate when you made your observations?	
8. Did you summarize and evaluate your experimental procedure, making comments about its success?	
9. Did you record if your hypothesis was true or false?	

10. If false, did you develop a new hypothesis that can be tested?	