Unit 1: Data Summary and Visualization I
2. Data basics

Stat 140 - 02

Mount Holyoke College

THE MATH/STAT DEPT
VIRTUAL TEA SERIES
PRESENTS

Open House

2020

COME JOIN US

08/27 Thur,
3:15PM-4:30PM

Email Sheila Heady to register
srheady@mtholyoke.edu

WHAT TO EXPECT

- Meet your fellow math/stat students
- Meet your professors in a fun and stress-less way
- Lear about the resources and opportunites at the department

PROGRAM HIGHLIGHT

Student summer research

Graduate school application

Writers in math/stat workshop

Study abroad info session

Course advising

Department of Mathematics & Statistics

1. **Data** are values, whether numerical or labels, together with their context.
2. **Observational units** are what you take measurements on
3. **Variables** are the characteristics recorded about each individual
4. **Categorical variables** identify a category for each case. They have a limited number of different values, called **levels**. E.g., Maritial status is a categorical variable, and the levels are single, married, divorced, widower, etc.
5. **Quantitative variables** record measurements or amounts of something

Suppose we have a data set that consists of patients who entered the emergency room at French Hospital in the previous week.

Poll question

What are the observational units?

1. All patients in France
2. The French hospital
3. Patients who entered the emergency room in the previous week

Suppose we have a data set that consists of patients who entered the emergency room at French Hospital in the previous week.

Poll question

Indicate whether the following is a categorical variable, a numerical variable, or not a variable with regard to these observational units:

How long the patient waits to be seen by a medical professional

1. Categorical
2. Numerical
3. Not a variable

Suppose we have a data set that consists of patients who entered the emergency room at French Hospital in the previous week.

Poll question

Indicate whether the following is a categorical variable, a numerical variable, or not a variable with regard to these observational units:
Whether or not the patient has health insurance

1. Categorical
2. Numerical
3. Not a variable

Suppose we have a data set that consists of patients who entered the emergency room at French Hospital in the previous week.

Poll question

Indicate whether the following is a categorical variable, a numerical variable, or not a variable with regard to these observational units:
Day of the week on which the patient arrives

1. Categorical
2. Numerical
3. Not a variable

Suppose we have a data set that consists of patients who entered the emergency room at French Hospital in the previous week.

Poll question

Indicate whether the following is a categorical variable, a numerical variable, or not a variable with regard to these observational units:

Average wait time before the patient is seen by a medical professional

1. Categorical
2. Numerical
3. Not a variable

Suppose we have a data set that consists of patients who entered the emergency room at French Hospital in the previous week.

Poll question

Indicate whether the following is a categorical variable, a numerical variable, or not a variable with regard to these observational units:

Whether or not wait times tend to be longer on weekends than weekdays

1. Categorical
2. Numerical
3. Not a variable

Suppose we have a data set that consists of patients who entered the emergency room at French Hospital in the previous week.

Poll question

Indicate whether the following is a categorical variable, a numerical variable, or not a variable with regard to these observational units:
Total cost of the emergency room visit

1. Categorical
2. Numerical
3. Not a variable

- ▶ What is the research question?
- ▶ What is the population of interest?
- ▶ What are the observational units?
- ▶ Name all the variables.
- ▶ Specify for each variable whether its use indicates that it should be treated as categorical or quantitative.

Tutorial exercise: 10 minutes

Finish Topic 1 part 2: online shopping
See link in zoom chat or on daily schedule
Goal: practice identifying observational units, categorical variables and numerical variables
Type your questions in the chat, if there are any
Put 'raise your hand' button on once you are done
Complete Topic 1 part 1.

Data is usually represented by a data matrix
- ▶ row: observational units
- ▶ column: variables

year <int>	month <int>	day <int>	dep_time <int>	dep_delay <dbl>	arr_time <int>
2013	6	30	940	15	1216
2013	5	7	1657	-3	2104
2013	12	8	859	-1	1238
2013	5	14	1841	-4	2122
2013	7	21	1102	-3	1230
2013	1	1	1817	-3	2008

Some helpful R commands to have a first look of your data matrix

- ▶ head
- ▶ str
- ▶ dim
- ▶ nrow (or ncol)
- ▶ names
- ▶ $

Tutorial exercise: 15 minutes

Finish topic 2 and topic 3
Goal: practice using R command for data matrix/frame
Type your questions in the chat
Put 'raise your hand' button on once you are done
For the rest of the time, work on 'reading a news activity' on piazza.

1. Identify the 5W's
2. Understanding the data table