Introduction to Art of Stat
2021-22
Numerical Summaries

www.artofstat.com

Before You Begin:

__#1. Right click to download from the course website: ***lung_demo.xlsx***

__#2. Launch Excel. Open lung_demo.xlsx. Minimize Excel. Do NOT exit.

__#3. Launch ArtofStat at http://www.artofstat.com/

Key: lung_demo.xlsx

This excel file has 4 sheets; the tabs to access these are at the bottom of your screen

area
area and sex
area and mfvc
source

Introduction
UCLA Study of Chronic Obstructive Respiratory Disease

The lung function data for illustration is a subset of the data from UCLA study of chronic obstructive respiratory disease (CORD). The original study followed over 15,000 persons and obtained measurements of lung function (FVC and FEV1, explained below) at two points in time so that they could investigate the change in lung function in relationship to location of residence, a proxy for exposure to air pollution.

The data used for this illustration (**lung_demo.xlsx**) is for the first time period only. It is a sample of n=150 ***families.*** There are 6 variables.

References:
Detels R., Coulson A, Tashkin D and Rokaw S (1975). Reliability of plethysmography, the single breath test and spirometry in population studies. *Bulletin de Physiopathologie Respiratoire*, **11**, 9-30.

Tashkin DP, Clark VA, Simmons M, Reems C, Coulson AH, Bourque LB, Sayre JW, Detels R and Rokaw S (1984). The UCLA population studies of chronic obstructive respiratory disease. VII. Relationship between parents smoking and children's lung function. *American Review of Respiratory Disease*, **129**, 891-97.

Key:

FVC = Forced Vital Capacity (liters). It is the amount of air that can be forcibly exhaled after taking the deepest breath possible.

FEV1 = Forced Expiratory Volume in 1 Second (liters). It is the amount of air that can be forcibly exhaled during the first second of an FVC test.

Coding Manual

Variable Name	Coding	Description
id	1 to 150	Identification Number
area	1 = Burbank 2 = Lancaster 3 = Long Beach 4 = Glendora	
mfvc	Continuous, liters	FVC, mother (liters)
mfev1	Continuous, liters	FEV1, mother (liters)
ocsex	1 = Male 2 = Female	Sex, oldest child
ocfvc	Continuous, liters	FVC, oldest child (liters)

1.
Single Variable Descriptives

1a. Discrete Variable: area

Excel file used: lung_demo.xlsx
Sheet used: area

You will do an EDIT/COPY from Excel followed by a PASTE into Art of Stat.

In Excel

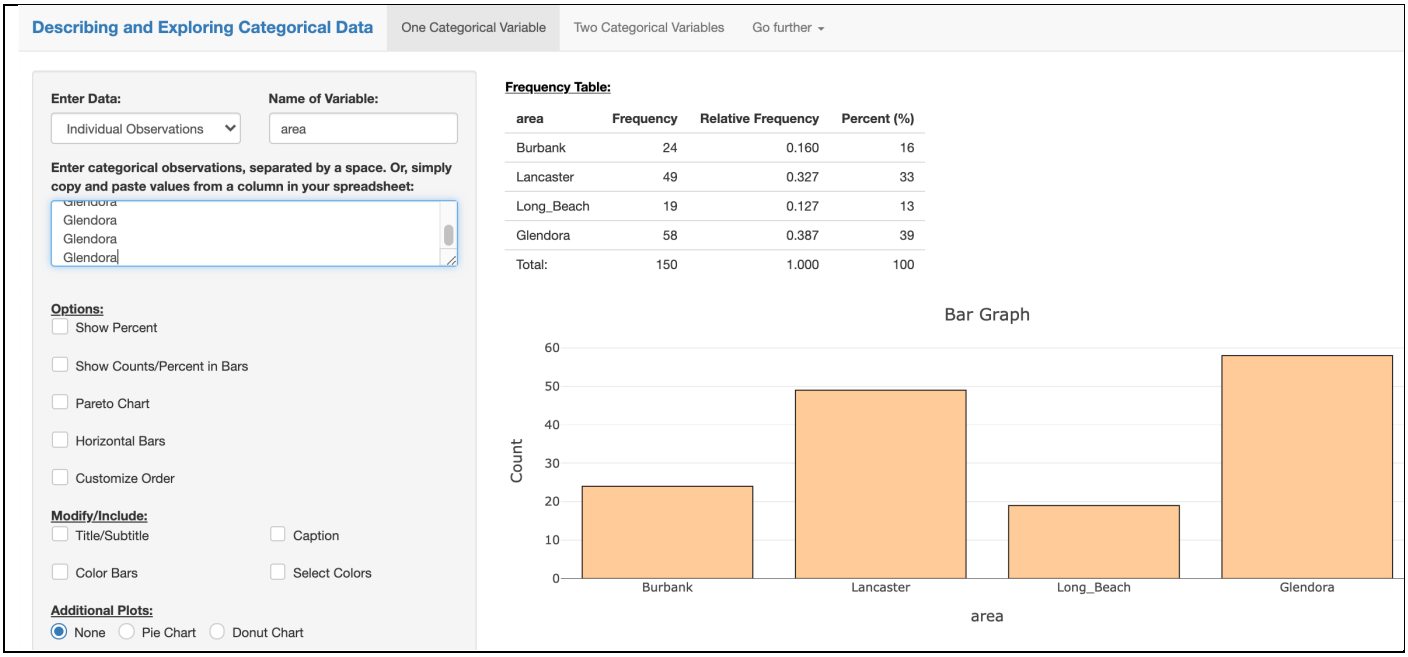
At bottom click on tab for sheet = "**area**"
Drag your cursor to select all 150 rows of data
From top menu bar EDIT > COPY
Minimize Excel but do not exit

In Art of Stat

At top right, click: **Online Web Apps**
At top left, click: **Explore Categorical Data**
At top, choose tab: **One Categorical Variable**

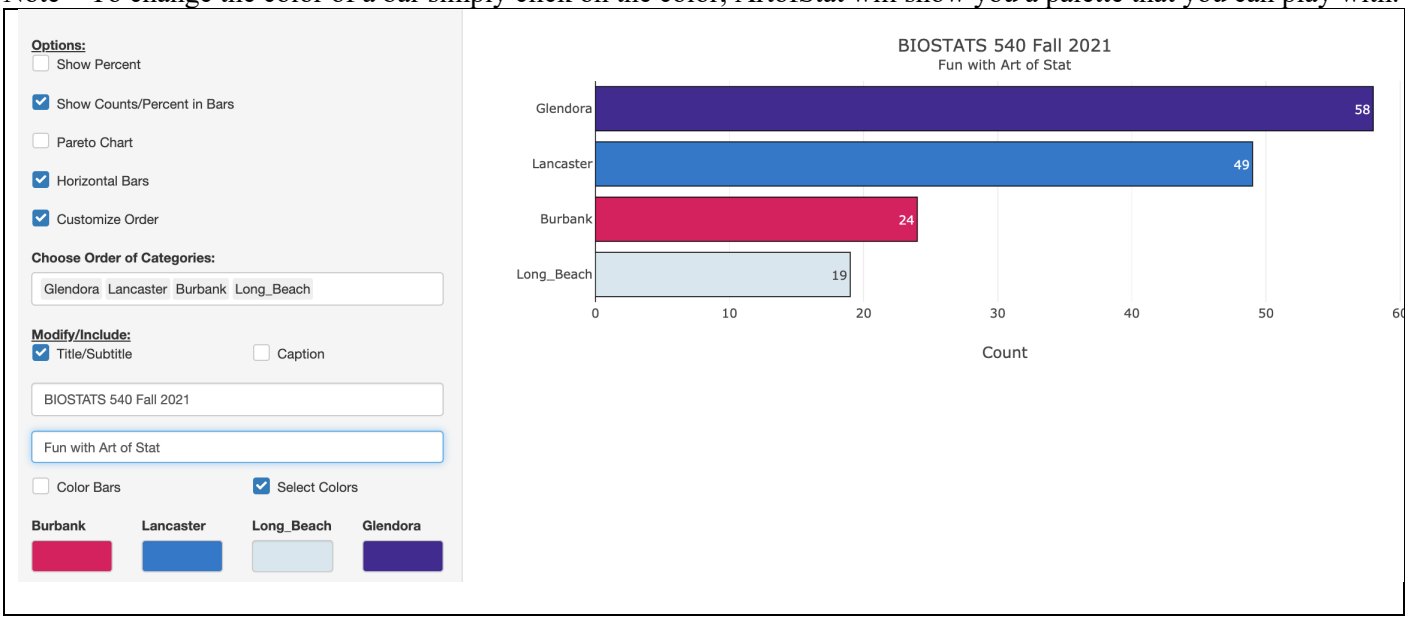
Now take a look at all your options on the left (you can play with these later!)
At Enter Data, choose: **Individual Observations**
At Name of Variable, type: **area**
Place your cursor in the box that reads "Enter categorical observations, separated by …"
From top menu bar EDIT > PASTE (alternatively you could do cntrl-V (Windows) or command-V (Mac))

You should now see:



Key:
In the frequency table we see that the total sample size is 150.
There are 24 observations of "Burbank", 49 observations of "Lancaster", etc
These counts are plotted in the Bar Graph

Next (just for fun), play with the options at left.
E.g, I clicked on "**Show Counts/Percent in Bars**", "**Horizontal Bars**", "**Customize Order"**, and Changed the title.
Note – To change the color of a bar simply click on the color; ArtofStat will show you a palette that you can play with.



1b. Continuous Variable: mfvc

Excel file used: lung_demo.xlsx
Sheet used: area and mfvc

In Excel

At bottom click on tab for sheet = "**area and mfvc**"
Drag your cursor to select all 150 rows of mfvc data of the second column, starting with the 2nd ROW
Tip – Be sure NOT to include the first row in your "edit/copy"
From top menu bar EDIT > COPY
Minimize Excel but do not exit

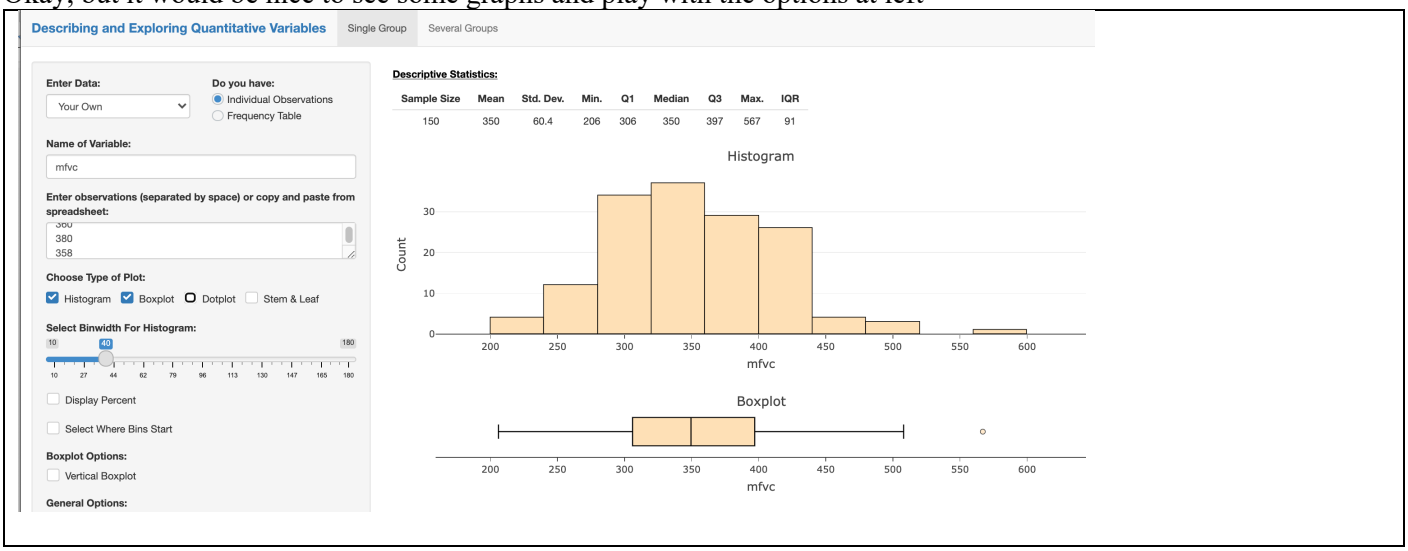
In Art of Stat

At top right, click: **Online Web Apps**
At top left, click: **Explore Quantitative Data**
At top, choose tab: **Single Group**

From the options at left
At Enter Data, choose: **Your own**
At Do You Have, choose: **Individual Observations**
At Name of Variable, type: **mfvc**
Place your cursor in the box that reads "Enter observations, separated by …"
From top menu bar EDIT > PASTE (alternatively you could do cntrl-V (Windows) or command-V (Mac)

You should now see:

Okay, but it would be nice to see some graphs and play with the options at left



Great hack: Hover your cursor over each plot. ArtofStat rewards you with information!
Play with the options at left.
E.g., I clicked **Histogram** (and set the binwidth = 40), and clicked **Boxplot**

Don't worry if you don't know the meaning of what you are looking at. Stay tuned!
The table of descriptive statistics shows you the values of all kinds of things that you will be learning in Unit 1.
The graphs will be explained in Unit 2

2.
Multiple Variable Descriptives

2a. Two Discrete Variables: area and sex

Excel file used: lung_demo.xlsx
Sheet used: area and sex

You will do an EDIT/COPY from Excel followed by a PASTE into Art of Stat.

In Excel

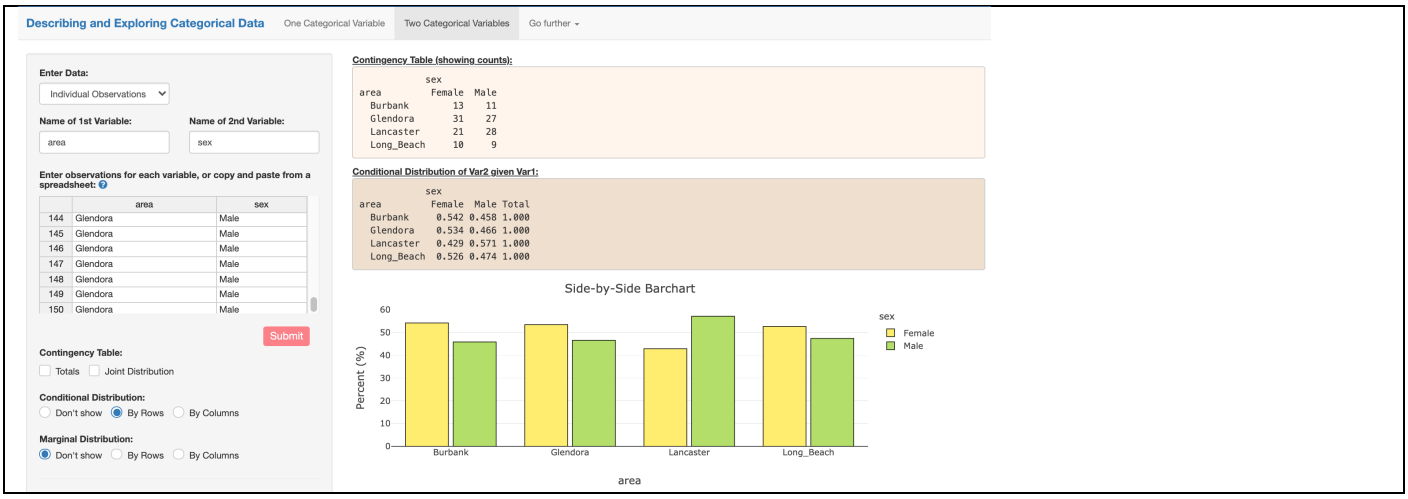
At bottom click on tab for sheet = "**area and sex**"
Beginning with row 2 (TIP – do NOT include row 1 as this contains the variable names, not data)
Drag your cursor to select all 150 rows of data
From top menu bar EDIT > COPY
Minimize Excel but do not exit

In Art of Stat

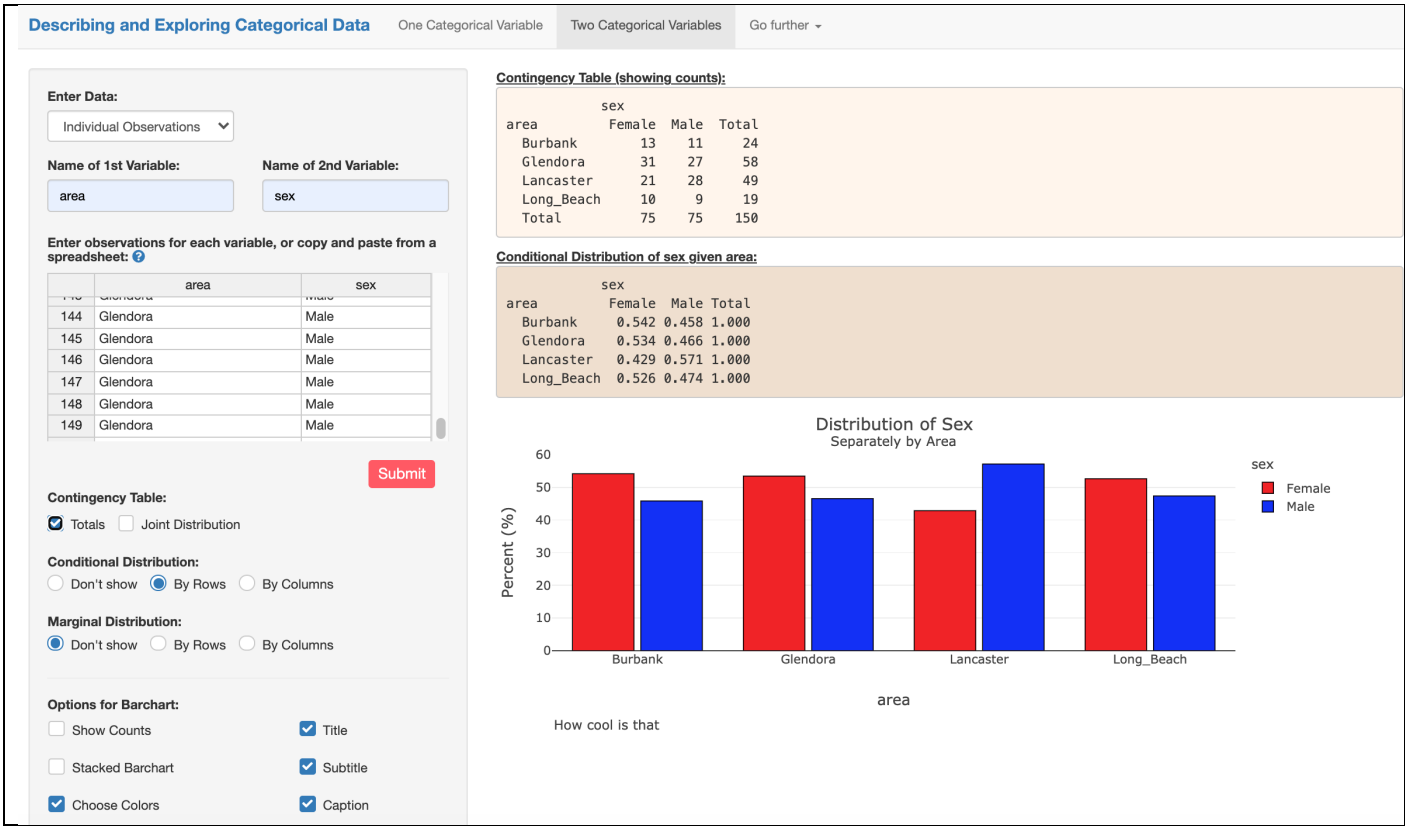
At top right, click: **Online Web Apps**
At top left, click: **Explore Categorical Data**
At top, choose tab: **Two Categorical Variables**

Now take a look at all your options on the left (you can play with these later!)
At Enter Data, choose: **Individual Observations**
At Name of 1st variable, type: **area**
At Name of 2nd variable: type: **sex**
Place your cursor in the first cell (row 1 column 1) of box that reads "Enter observations for each variable…"
From top menu bar EDIT > PASTE (alternatively you could do cntrl-V (Windows) or command-V (Mac))
Click on the red box **SUBMIT**

You should now see:



Here, I definitely want to change what is displayed and how it is displayed. Again, I played with the options at left:



Key!

Again. Please don't worry if you don't understand everything that you are looking at here. We will get to it! What's shown in the graph is the following. For each area separately, we can see the percent of repondents whose sex at birth is female v male.

2b. One Continuous (mfvc) by Group (area)

Excel file used: lung_demo.xlsx
Sheet used: area and mfvc

In Excel

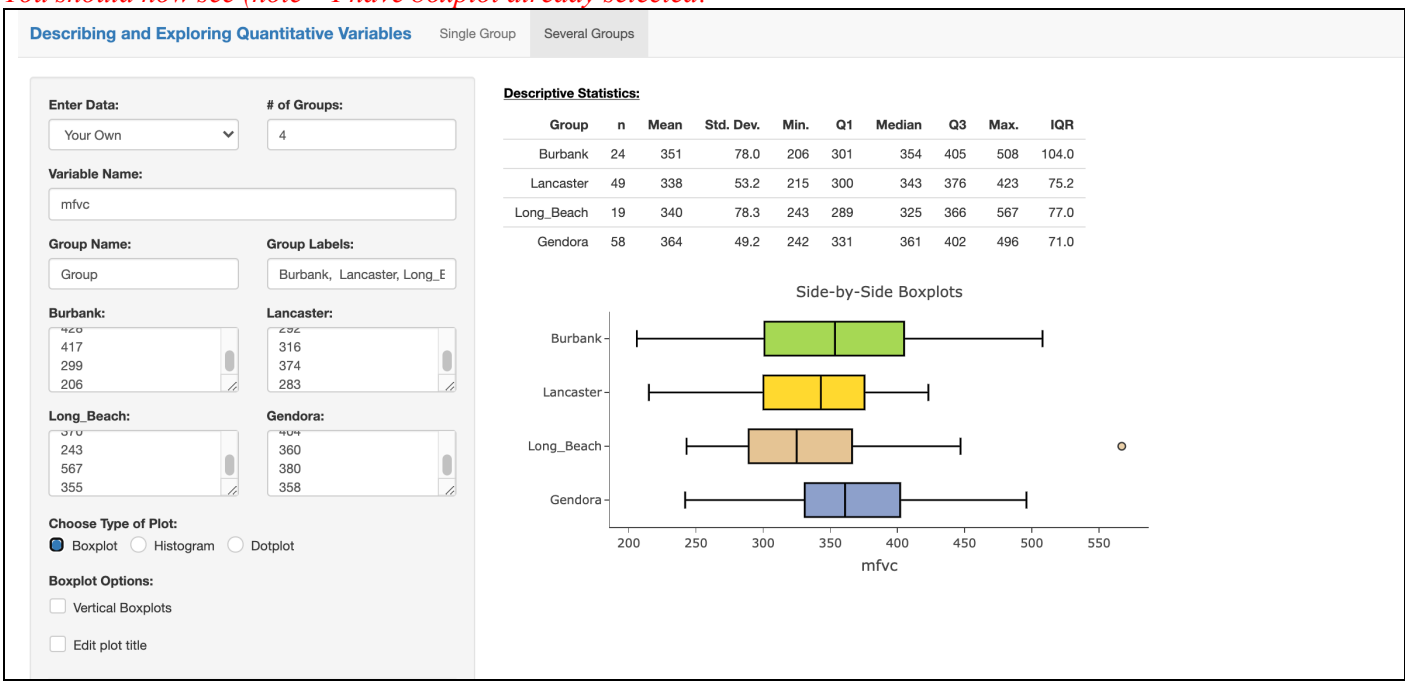
At bottom click on tab for sheet = "**area and mfvc**"
Beginning with row 2 (TIP – do NOT include row 1 as this contains the variable names, not data)
Drag your cursor to select all 150 rows of data
From top menu bar EDIT > COPY
Minimize Excel but do not exit

In Art of Stat

At top right, click: **Online Web Apps**
At top left, click: **Exploring Quantitative Data**
At top, choose tab: **Several Groups**

Now take a look at all your options on the left (you can play with these later!)
At Enter Data, choose: **Your Own**
At Number of Groups: **4**
At Group Name: **area**
At Group Labels: **Burbank, Lancaster, Long_Beach, Glendora**
At this point you will have to do 4 separate EDIT>COPY>PASTE of the values of mfvc, one for each area
From top menu bar EDIT > PASTE (alternatively you could do cntrl-V (Windows) or command-V (Mac))
Click on the red box **SUBMIT**

You should now see (note – I have boxplot already selected:



Following is what I get after playing with the options at left:

Describing and Exploring Quantitative Variables Single Group | Several Groups

Enter Data: Your Own

# of Groups: 4

Variable Name: mfvc

Group Name: Group

Group Labels: Burbank, Lancaster, Long_I

Burbank: 417, 299, 206

Lancaster: 316, 374, 283

Long_Beach: 243, 567, 355

Gendora: 360, 380, 358

Choose Type of Plot: (•) Boxplot ( ) Histogram ( ) Dotplot

Boxplot Options:

[x] Vertical Boxplots

[x] Edit plot title

Histogram: Side-by-Side H

Boxplot: Distribution of M

Dotplot: Side-by-Side D

Descriptive Statistics:

Group	n	Mean	Std. Dev.	Min.	Q1	Median	Q3	Max.	IQR
Burbank	24	351	78.0	206	301	354	405	508	104.0
Lancaster	49	338	53.2	215	300	343	376	423	75.2
Long_Beach	19	340	78.3	243	289	325	366	567	77.0
Gendora	58	364	49.2	242	331	361	402	496	71.0