Data Visualization Using an Open Source Statistical Program - RStudio

Author: Shaley Valentine

Grade Level: $10^{th} - 12^{th}$ grade

Group Size: 1-2 per station; full class

Setting: Computer Lab: Program installed prior to lesson

Time Needed: 2 x 50 minute sessions

Equipment Needed: Computer; Installed version of R/Rstudio (most recent; installation instructions in lesson)

Objectives

Composite Learning Objective: Students will gain a basic understanding of R statistical software and develop basic statistical and graphing skills using R. Students will use data collected by researchers to (a) compute and visualize simple summary statistics comparing groups of Lake Sturgeon; and (b) assess statistical relationships between different variables.

Knowledge Outcomes:

- Students will learn how to install and navigate the R/Rstudio environment.
- Students will learn how to use basic code input to conduct simple data visualization and computation.

Skills Outcomes

- Students will compute and visualize summary statistics for comparing two or more groups by their sample average (μ).
- Students will use R/Rstudio to assess statistical relationships between different variables.

Disposition Outcomes:

- Student will have an adequate understanding of how fisheries professionals use R/Rstudio to visualize different fish population characteristics using a threatened fish species to provide context.
- Students will understand the importance of open source programming to answer basic fisheries questions.

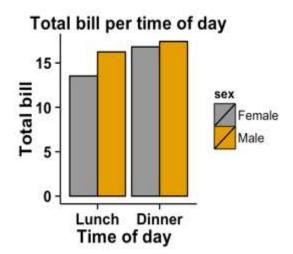
Summary

R is a free, open-sourced statistics program that is widely used by scientists. Statistical analyses, summary statistics, and data visualizations are important tools for students to learn and use. These tools help to simplify relationships or differences between variables. Students will go through a tutorial using program R to learn how to conduct basic summary statistics and analyses and will use graphing functions to compare groups. Students will use data collected by MSU/MiDNR Black Lake Stream-side Facility researchers to evaluate relationships between different variables.

Background

R is free statistical software that is used widely by scientists. It is free to download from the internet and all the code is open-sourced, meaning that you can look at how people created the analyses and replicate that code. A simple way of explaining code is that it is a language. In the case of R, the language the code-writer speaks is interpreted, rather than compiled. Simply put, this means that what you type is interpreted by the computer as a command, rather than as a program. So if you tell the computer to conduct an analysis, that's what it does. Whereas an application on your iPhone is a program, which requires each request to be confirmed by a coded program. As a result, R is intuitive and easy to reproduce. Additionally, other programs like SPSS, SAS, or C++ are extremely costly and generally proprietary. As such it may be difficult to replicate analyses conducted by other researchers.

We can use R to calculate **summary statistics** including the **sample mean**, **median**, **mode**, **quantiles**, and **standard deviation**. These statistics can tell us general trends in data and among groups within data. For example, we could look at average height between males and females to get a general idea of differences between sexes (Figure 1 below). What we see is that females appear to spend less money both on lunch and dinner, and both males and females tend to spend less on lunch than dinner. Graphs like Figure 1 and summary statistics allow researchers to make generalizations about data and compare one or more groups to each other. We can also use data visualization to evaluate other relationships including those which are **predictive**. Figure 2 shows an example of a predictive relationship, where a **scatterplot** is used to show the relationship between height (in) and speed (mi/hour). In this case, we can use data visualization to infer how the value of a predictor variable can affect the magnitude of a **dependent variable**.



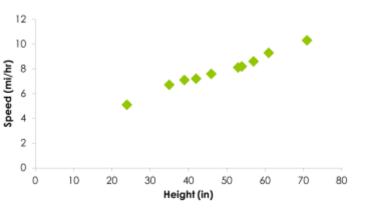


Figure 2. Walking speed (mi/hr) as a function of height (in)

Figure 1. Cost of lunch organized by sex and time of day

Researchers use **Statistical analyses** to evaluate whether differences in data are "significant." For example, the mean cost of lunch may appear to be different between males and females in Figure 1, but this may not be meaningfully different based on a statistical analysis. Simple analyses include like correlations (looking at how variables are related to each other) and T-tests (comparing the mean values of two groups). Analyses can be powerful, but they do have downfalls. For example, correlations show how related two variables are to one another. Often, two unrelated variables may be correlated, but that correlation may not be of consequence. For example, the price of pickles may increase as your age increases. While this may be true, it's not your age that causes the price of pickles to increase, thus it's important to remember that correlation does not equal causation; and that all statistics must be used with care.

In this lesson, we will use RStudio to visually and analyze data collected by researchers at the Black River Streamside Rearing Facility, where researchers have studied threatened Lake Sturgeon since 2001.

Definitions

Summary statistics: information that gives a quick and simple description of data. Can include mean, median, mode, minimum value, maximum value, range, standard deviation.

Sample mean: the average of n observations from the sample. The sum of all values in a sample divided by the number of values in a sample.

Dependent variable: a variable (often denoted by y) whose value depends on that of another.

Median: The midpoint value of any sample. Where the number of observations is an even number, the median is the average of the two midpoint values.

Mode: The value in a sample population which appears most frequently.

Quantile: A subset of a sample divided into equal groups. If a population has values of 1, 2, 3, 4, 5, and six, a quantile might represent 2 units; ie: [1,2], [3,4], [4,5].

Predictive (independent variable): The process by which a value can be used to infer the magnitude of another value which directly results from the first. (a variable (often denoted by x) whose variation does not depend on that of another).

Scatter plot: a graph in which the values of two variables are plotted along two axes, the pattern of the resulting points revealing any correlation present.

Standard deviation: A value used to quantify the amount of variation of individual observations from the sample mean. Ie: One might have a sample of fish sizes: 130cm, 135 cm, 140cm and a second sample of fish sizes: 130cm, 130cm, and 145cm. Both groups have a mean of 135cm. Standard deviation allows us to determine how much each individual differs from the sample mean. In this case, group 2 deviates 8.66cm from the mean, where group 1 deviates 5cm from the mean.

Statistical analyses: mathematical comparison between one or more discrete groups and their variability.

Activity

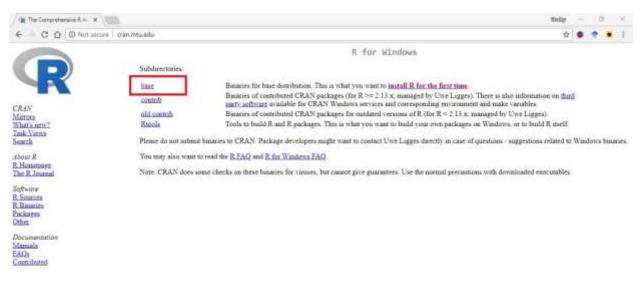
Download R

First, we need to download R. Go to the [cran-R website] (http://cran.mtu.edu/) and under "Download and Install R" click on the link that your computer is. After clicking on the link you will need to click on "base" under "Subdirectories" in the next window. In the next window click "Download R 3.x.x for Computer Type". Follow the images below for help:

Step 1: Download and Install R

C O @ Not secure crans	mtuedu 🖞	4		÷.
	The Comprehensive R Archive Network			i
CRAN Minutas What's new? Task Views Search About R R.Homepage The R.Journal Software R.Sanacas R.Sanacas R.Sanacas Packares Other Documentation Marcelia EAOs Conscibuted	Described and Install # Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of Reserversions of R. Download R for Lature Download R for Lature Download R for MactOS X Download R for MactOS X Download R for MactOS X Download R for all Platforms R is put of many Lature distributions, you mould check with your Linux package management system in addition to the link above. Sources Code for all Platforms Windows and Mac users more likely want to download the precompiled binaries listed in the upper box, not the source code. The sources laws to be compiled before you can use them. If you do not know what this means, you probably do not want to do it! The latest telesse (2018-07-92, Franker Spray) <u>R-3.5.1 tar.pr</u> read yhati aren in the latest version. Sources of <u>R alpha and heta releases</u> (analy unaphots, created only in time periods before a plauned release). Download estemation of R is available here. Courributed estemation of R is available here. Courributed estemation of R is available here. Plays there a flow of the precision about R like how to download and install the software, or what the license terms are, please read our armits to Disparently unkel quantions before you send an email.			

Step 2: Click on "Base"



Step 3: Select "Download R 3.x.x for Windows

← C ① @ Net secure c	animtuledu	ά.	•	÷	• 19
	R-3.5.1 for Windows (32/64 bit)				
	Download R. 3.5.1 for Windows (62 megabyes, 32:64 bit) Installation and other instructions lists feelines in this version				
What's new? Task Menus Search	If you want to double-check that the parkage you have downloaded matches the package distributed by CRAN, you can compare the <u>unifourn</u> of <u>fingerprint</u> on the master server. You will need a version of md5sum for windows: both <u>graphical</u> and <u>command line versions</u> are available. Frequently subst questions	the	est l	io the	
Abour R R. Homepage The R. Journal	Dates R can under not version of Windows? How do Luedare markages in not previous version of R? Should I run 32-bit or 64-bit R?				
Software R. Sinetans R. Binaries Packastis Other	Please see the <u>R.FAQ</u> for general information about R and the <u>R.Windows FAQ</u> for Windows-specific information. Other builds				
Documentation Manuala EAQs	A build of the development version (which will eventually become the next major release of R) is available in the <u>side of analytic build</u> . Previous releases Note in webmastery: A stable link which will reduce to the current Windows binary release is				
Concidentia	Note in webmasters: A stable into which will reduce to the current Window's binary release is «CRAN MIRROR» hin windows base release htm Last channe: 2018-07-02				

You will also need to download R studio. Go to the [download

site](https://www.rstudio.com/products/rstudio/download/#download) and scroll down to "Installers for Supported Platforms" under "RStudio Desktop 1.x.xxx". Choose the download based on the computer you have.

Step 4: Download "R Studio"

Downiceti Ri	Budo Bi. × UIII					80	lig -	.0	
+ c 0	Secure http://www.studio.com/products/triudio/download/*utranicad						*	. * 1	•
	R Studio	Products	Bestown	ces Pricing	AboutUs	Blogs Q			
	RStudio requires R 3.0.1+. If you don't already have R, downi	oad it here.							
	Linux users may need to import RStudio's public code-signin policy.	ng kary prior to	installation,)	depending on the	e operating system	n's security			
	Installers for Supported Platforms	-							
	Installers	Size	Clate	1/05					
	Retructio 1, 1, 401 - Windows Vista (1/0/10)	35.8715	2018-07-19	24ca3fe8dad81	87.aabd4bfbb9dc2	boet			
	R25u01011408-1Mac 05X104+104-010	743105	2018-07-19	4fc4f4f78845b	142hf96dcla5h1d	c556			
	Rituano 11409 - Liburru 1204-1510, Detram 5132 (cd)	把工机 器	2018-07-19	349349d5839e3	a34697748676820	6961			
	R35.idio 11.455 - Liburtu 11.04-15.10/Deblari 5.84-00	97.4 MB	1018-07-19	8634488612035	8Fa0145o4d14cd7	50.64			
	Ritudio 1.1 455 - Ukumu 16.04+ (Dahan 8+ (64-b.))	64,3148	1018-07-19	d96e63548c2ad	d8906ac633bd688	3432			
	Riturio 11 455 - Ferrira Div/RetHat 74 (speciality 11 14 (1044))	101108	2018-07-19	1#f56c7cd80e2	634fBa9fddiical	612.0			
	Riskatio 11,458 - Fedura 19-/Rechtst P-/Ispen5USE 111- (64-bit)	90446	2018-07-19	5e77094a88fdb	ddddadd3579875	2462			
	Zip/Tarballs								
	Zip/tar archives	529	Date	MOS	Try RSh	udio Server Pro for f	freel	-	
	#55.adio 11.458 - Windows Mata (7/0-10	122 年14日	1018-07-18	655d6b+e716d8	c97acb/50127009	9613	_	-	
	When we had not the second second by the second second second	100 March 100 Ma	Contra and con	and the second of					

Once both R and R studio are downloaded from the website, follow their prompts to install them on the computer.

NOTE: When you use R, open R studio only. You do not need to open Base R as well. To use a simple analogy, R and RStudio work together like the motor and dashboard of a car. You can certainly get to point A with an engine, but the process is a lot easier if you can use a dashboard user interface. Working in R, one would only enter their code. One error can result in code failure and the loss of hours of work. RStudio lets the user revisit and generate code, is much more intuitive and visually pleasing, and allows for simple modifications to code, rather than an "all or nothing" approach.

###Windows in R

When you open R studio, you will see 4 windows. This section explains the purpose of each.

101-0	- Code You From Section Build Coding Paths Sam Many						-	σ	×
9. * =	z + G Gl 🖂 a senstantenne 🗌 🗐 + Adem -							Project	
-	neelin = 🖭 untreelin = 🔍 Assoc Analyses And = 🖓 Tertilabilitan Analysis R =	P Riesson pron Rind =	Multiple Tabs	tre	monutert His	tory .	Environm	ent	-01
	2 1 7 9 a mon . 0 .	10	ned + 9 0 _+Pun + 4- 2	1.00	16 Pmp	+ Hutted +	1	There -	1.6
37	eewindus in a		· · · · · · · · · · · · · · · · · · ·	15	Global Environm	rent +	(Q.)		
38	So you see a lot of windows open in a studio.	iding Window		Dat	CONTRACTOR OF THE OWNER		12 M 11 M 12 M		
40				11.77.7	ser, hatch		of 8 variables		
	 This window (top left) is the coding window. You type is multiple tabs open in this window. 	ode into different d	scuments, and you can have		er, sort		of 8 variables		.문.
- 42				11. TO 3			of 6 variables		10
	-The window below (buttom left) is the Console. This is you run it.	where code is shown	and warnings and whatnot after		utal		of 10 variables		
- 44	which is present the second	an ann ann ann ann ann ann	and the second				of 10 variables		10
	 The window in the top right is the anvironment. Any dat are shown here with the sommary of the data and objects click on the history tail as well to see what functions y 	(size, observations,		100					
46	-The window in the buttom right has a lot of tabs on it, opened in E, and you can click on them to bring them in Figures you made in R. The Packages tab will also you do currently in use. This is your "fibrary", the belp tab	The Plots tab will hat packages you have is a way to search for	they you plots and other installed and which ones are	1.54	ewer	ckapes Halp	ackages,	He	PA
46	search will pop up here. I don't know what the viewer to	NP 16-			Hane	Description		Wini-	
49	AANTIN ANAL			Une	Library				-
	C Drevitad # 5		R Markitown 5		shind	Contribution	Utalimentional Acapt	14.5	0.77
Come	er et al				acepack	Regeniiio	käl kur Serectssig Wultipre Riemethermetherne	3,43	0
Туря	Ticense()' or 'Ticense()' for distribution details.		~ C		NUMBER		FreePress A 120 F	8.2.8	0
4,15	a collaborative project with many contributors. contributors() for more information and	Console			backports		etations of Fanctions Since 8-3.3.3	54.5	0
CITR	tion()' on how to cite # or # packages in publications.				baseMenc	Tank for to	celil exceding	3.1-1	0
					betareg	Deta Negres		-114	0
	'demo()' for some demos, 'help()' for un-live help, or			11	84	Boart C++			
'belp.	start()' for an nTHL browser interface to help. 'u()' to guit 8.			1.00				1825-	
'belp. Type	'ųQ° το quit π.				konde		d Advellentergs	1828- 1 0.11	0
'help. Type					kinair bineircate			1825-	

- The top left window is the Coding Window. You type code into different documents, and you can have multiple tabs open in this window. If you've written code that you think will be useful to revisit for future projects, this window can be saved and reopened. In the above example, the author has five windows, or "Multiple Tabs," open. Clearly, this author is working on a few different projects.
- 2. The bottom left window is the **Console**. This is where RStudio conveys the results of the code written in the **Coding Window.** Here, the user can see specific error messages, and can ask RStudio for help if needed using the command 'help()'.
- 3. The window in the top right is the **Environment**. Any data or objects you have read into or created in Rstudio are shown here with the summary of the data and objects (size, observations, variables, etc.). You can click on the history tab as well to see what functions you used recently. More often than not, the user does not use this window.
- 4. The window in the bottom right has a lot of tabs on it. The files tab will show you files you recently opened in R, and you can click on them to bring them in. The Plots tab will show you plots and other figures you made in R. The Packages tab will show you what packages you have installed and which ones are currently in use. This is your *library*. The help tab is a way to search for help or anything you do search will pop up here. I don't know what the Viewer tab is.

###Dataset

We will be working with a data set of adult lake sturgeon captures from 2018. These are lake sturgeon that were handled by MSU/MiDNR researchers in May 2018. Length (centimeters; total and fork; see figure below), mass (kilogram), and girth (centimeters; diameter of fish) measurements along with gender identification are recorded in the data set.

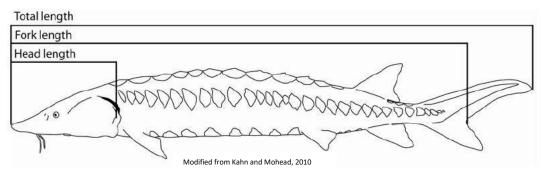
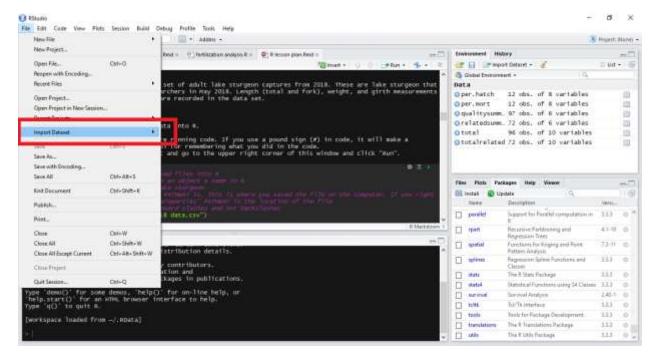


Figure 3. Total and fork length (cm) measurements on an adult Atlantic Sturgeon.

###Importing Data

First, we need to import data into R. Importing data means R will take a file on your computer and read it into the program so it is usable within R. You can find the dataset for this lesson via: <u>https://www.glsturgeon.com/sturgeon/behavioral-ecology/lesson-14-Introduction-to-R/</u>. The data is in an excel file, which you can download at the bottom of the page. Download the data to the desktop of your computer.



1. We need to import our data into R first. Go to File -> Import Dataset -> From CSV

 This next window will open once you click on Import Data from CSV. Click on the "Browse" Button.

1	(e.) Lorisber(? =	Red = ?)Terrispier.a	nitysta R. H.	P. R WISSIN DAY	Prod =			⇒C)	transact H	story	i int	
100	NiaMit									Denta-		and the second se
TO												
											00m. 113 13-10	1
0	import Options						-	e Preview		1	- 13	
	Norrie: dataset		Defecter	Comme -	Iscape	New Y		lifary(readr) taset <- read	csv(MAL)	1	11) 15-10 73-11	
2) 11c		Trim Spaces	Definition	Correra .*	Comment	Delault •	4	Intany(meader)	,cev(Matt)		13-10 13-10	
C) The Con Con Clon	Norrie: dataset		610.0	- External	Comment	and the second	4	lifary(readr) taset <- read	,CSV(MALL)	1	113 13-10 73-11 833 133	
2) 11c	Norrie: dataset	Trim Spaces	Guotesi	telvet -	Comment	Delault •	4	lifary(readr) taset <- read	.csv(1651.)	inger Care	111) 13-10 13-11 133 133 133 133 133 133 133 133 1	
Gi Line con tion tion tion tion tion tion	Norrie: dataset	Trim Spaces	Guotesi	telvet -	Comment	Delault •	4	lifary(readr) taset <- read	_ctiv(Mitt)		113) 13-10 13-11 133 133 133 1434	

3. The file explorer will open. Find the data wherever you downloaded it, and then open it.

ntrustin = P (protect? =	Ovose File		×	ary.	-
Import feet Data	- + + + + Decitos + Elessons 2013 + Elesson	v 5 SeathTimer	e.	ADatuet - 🧃	II 647 + 1
Fieldst	Organise = New Autoer			ent • (9,	
Data Preview Inspirit Options	Thesis A harre Detection (), I A Final endyce A Codemic Public Ferfiliation Cross Figures	6/10/2018 244 PM F	inge your own.) Je folier Rimeark Dane E., MO File Spela Wee Gaca	343 obs. of 12 variables 12 obs. of 8 variables 13 obs. of 8 variables 37 obs. of 6 variables 72 obs. of 0 variables 96 obs. of 10 variables 172 obs. of 10 variables	
Nerne: Sittinet Store av Singe: 0 Otter Spaces Coger Data View	This PC			lages Hidy Wexwe Bala G Description	- Vans.
Code Preview:			and the second se	Curvising Multidimensional Arrays ACE and AMA3 for Selecting Multiple	343 0
[] Harary(readr) Hataart <- read_cav(Willi)	File game adult 2018 data			Regenzan Transformations	
view(dataset)		Open /	Cancel	Enty Pre-seul Pict Austriants Reinightermetations of Functions	145 0
and the second second		restoryte		Intraduced Save R-1.1.1	
			besch4ern:	Task for basel# encoding	33-3 1
8			betang BH	Beta Regression Boost Con Header Files	314 1
and the second	Wipot Cancel	netturyarror and the	in strike		4.000
sermal cov("minit 2018 since cov") # the - s	warm. In the same fullder an title decement		toner	Perometrized Active Bindings	41.1
"C://users/shaley valentine/besktop/# lessor	# 2018/H Lesson"		initrate	An Topp' Interface to Active Readings	8.2.2
			1 Initope	lithowe Operations	554 1

4. Before you hit "Import", make sure the "Import Options" section is filled out exactly as in the image below. Name the file "Is" (lake sturgeon). This step tells RStudio that you are importing a file which is in a comma separated format and that RStudio should read it in with the exact column names.

NUMBER PILNER	= @ Baser Analyses Prot = ?	Terristomen analysis R =	@1 Burnson plan Pa	OT A		11.10	woment His	tory		-
Import fest Data		Adam Anno 19 Cardao North	a second second		0.1		- 63	of Dataset + 🥳	10 M	it + 1
File/124						* 6	Global Environm	ent +	Q.	
S DECEN	ine/Desiton/Rilesons 2018/8 Lesson	And States and		Errore.		Da	ta			
Concernant inter	the provide Kindstern Grief of Carde	a esan co o anse co			t look at numbers.	0		583 obs. of 32 varia		
Data Review					0 200		per_hatch	12 obs. of 8 variabl		
2D Date	Then See	n n	Girth Weig	nt Location			per, mort	13 obs. of E variabl 97 obs. of E variabl		
(Hinger)" (Uherati			(integer)" doal		the provide		the second second second	72 obs. of 6 variabl		
G				×	Total Length of	1.1	total	96 obs. of 10 variab		
Previoung Brat 30 art	782.					0	totalrelate	d 72 obs. of 10 variab	les.	
Import Options:					1.00					
Norme: N	First Row as	Deleviter Comma	· i see	Name and	• 2 ×	18	ns Pipts Par	Ragen Help Viewer		
Skip: 0	Nertex	Gentee Default	Comment	Deladt -	THE REPORT OF	- 01	instal (D u)	pdata Q.S		
	Pillion Spaces	the second se		Detailt *			Date	Description	Week	1
	2 Open Data View	ver Locale: Configu	Main Park.	United T		Us	stitury			
Code Preview:				0	R Markdown	10	shind	Combine Multidimensional An		Юc:
In the second se						10	Hopsek:	ACE and AMULTUR Selecting Ma Regression Transformations	(tiple: 3.4.1	ł., .
adult_2018_date	read_cov("C:/mens/S	haloy valentice/De	sktop/A lesso	na 2018/9		10	inserthat	Easy Pre stul Post Accertants	3.2.4	
Lesson/adult 20 view(adult_201					Heetervieror in fil	* 0	baciperti	Reimpherspectations of Function Introduced Server P-13.3	n 133	1
						Ū	besefi4erec	Task for baselit encoding	21-1	1
01 <u></u>						0	belang	Beta Regression	3.14	Ε.
			- Migori	Cancel	directorypror in fil		SH	Boast Cox Hander Files	1.82	5
	MALE HER CONTRACTOR	and the state of the state	No. of Concession, Name			. 0	taner	Personalized Active Bridlings	41.1	1.1.1
ENGL						D	hindicate	An Topp' Interface to Active	8.2.2	1
CLOUSER: //ihaley >	alentine/besktop/# lesso	ns 2018/# Lesson"				1000		Headings		

5. After you import, you should see the file you imported in the **Coding Window.** At any time, you can view the data by clicking on the open tab, or by entering the command View(Is) in the coding window.

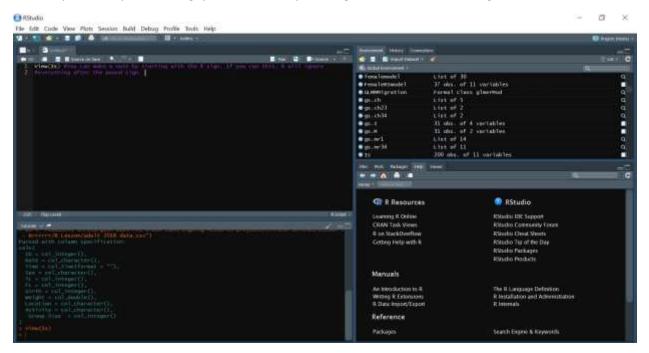
1 1	Othern								1 (m.)				10.00		0	
Image: Note:				Bernstell Hilling Commit												
Free type				C R. E worttweet-											- 16	
T VAUUE TRYING New 10 11 of 11 Life of 12 Ame owner 1 T VAUUE TRYING New 10 10 11 Life of 12 Ame owner 1 T VAUUE TRYING New 10 10 11 Life of 12 Ame owner 1 T VAUUE TRYING New 10 10 11 0 12 Ame owner 1 T VAUUE TRYING New 10 10 10 10 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 1 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 20 Ame owner 10 T VAUUE TRYING New 10 10 20 Ame owner 10 T VAUUE TRYING New 10	4	- 1/ 44		an Under Dertarment -	Terral Inc.	ALC: N	10.00									
2 XXXXXX Available 100 10 <td></td> <td></td> <td></td> <td></td> <td>the second se</td> <td>9.00020</td> <td>100 CAL</td> <td></td> <td></td> <td></td> <td>equine</td> <td></td> <td>eco</td> <td></td> <td></td> <td></td>					the second se	9.00020	100 CAL				equine		eco			
1 NVMM var NVMM NVMM NVMM <td></td>																
						distant in	III vien					Main		SECON		
2: 340000 City des 0 00 0						diseased.	ON NAME					- 1448		SWOOTH .		
2 - 3.4000 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0						queening	NJ (Mer	HI I				Malk		596/0114		а.
Image: 1 Support 1 Week Week 10 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10 Dot men week 10 Dot men week 10 Dot men week 10 Image: 1 Support 1 Week 10						and the second s	and the second					-		5,9,0111		
1 1 1000000000000000000000000000000000000						granting .	Kol sken			145		-	171608	5,9,0018		
A ANOTH AN AND THE DATA						(parents)	SOL COMP.				-	i inte	1222108	SMOOTH		
15 AVOID 1000 1000 100 100 100 100 100 100 100						and the second	test them		446			Main		LAUTE		
It Neveral transported in the second of t		of 11 variables	200 obs. of 11	• 16		and the second	HALE COMP					1400				
It Neveral transported in the second of t			Later .	Inc. inc. Description		Summing.	the own			145	-		104000	VALUE		
to Average of Average			PAARS									-	121100	STRONG B		
TE: Aquetta: transformedia ratio Te: en attraction researching to transformedia ratio Te: en attraction ratio Te: en attractractine attraction ratio Te: en attra																
It support times will be at the at the analysis of the analysis of the at the analysis of the at				Martin Contraction												
s he to the terms of the second secon		Disudia		Con the Designation of the												
Lawring & Correct State DX Appendix Constant DX App		- Bacablo		Ner Kinsburges												÷.,
R os stacktonfors Cathing Hiele with A Stacktonfors Cathing Hiele with A Stacktonfors Cathing Hiele with A Stackton Tool Hield y Stackton Packages Stackton Pack		RStatio TOC Sapport		Loarning R Online												
An end with a set to a set the set of the se		RStartio Community Future		CRAN Task Wews												
and with sultain spectfullities. Control tide with it is an interval to a subscription of the Day Riskdo Packages is a subscription of the Day Riskdo Packag		RStartio Cheat Sheets		R on StackCrawfave	ILLING ALL DRIVE TO A					NO	Concession of the	201	the set of	The second		
Kitada Products Kitada Products Kitada Products Kitada Products Kitada Products Kitada Products Marisalis Second Statements Kitada Products Marisalis An introduction to 8 Marisalis Marisalis Kitada Products		RStathe Tap of the Day		Cetting Help with B									1000	ullaters (
Kitada Products Kitada Products Kitada Products Kitada Products Kitada Products Kitada Products Marisalis Second Statements Kitada Products Marisalis An introduction to 8 Marisalis Marisalis Kitada Products																
en en ani- L'Anna Arren 1 - 112. en ani- L'Anna Arren Santa - 112. en ani- L'Anna - 112. ani- L'Anna - 12. ani- L'Anna - 12. ani- L'Anna - 12																
													Line .	A Long Lab		
- Levis, Version 2017, Version				Manuals									0.01			
anth - avit semander/s B De Eulergedo Defension Martender/s et al. Adventer/s B De Eulergedo Defension With a stat Adventer/s B Defension With a stat Adventer/s B Defension B Data Impact/Support B Defension B Data Impact/Support B Defension																
right a ruit dividie(). Withing & Schward and Administration and him - un Lateractive(). It Data import/Sport & Hermula triving - and Administration																
try try = in Linkerse for Ch.																
		2 Internals		II Data import/Export												
				Reference												
landing and an and a																

6. Before we can evaluate our data, we need to open a new R Script in the **Coding Window.** You can do this by selecting the paper and plus icon in the upper right corner, then select "R Script".

				100	ALC: NO						O right plan
Plane Clarbears									Assessment Annual Connects		
A Medition									C		
E Vietzzen.					-	-		1001	Es little Internet -		4
Trees Hale Ages	10122-04		e	1997		A DECEMBER	A GEORGE		fesslewate1	Alter of 30	
	10,000					U aler	Statements 1	1	Fena leksede 1	37 obs. of 11 variables	
C++-Im		Mate				12 dans	(prevente)		GLAMMigration	Formal class glmorted	5
		- Marine				e den	queting		49.05 18.073	List of 2	2
	- 44	Male				ut den	quining		as.ch34	List of 2	
	17111	1			#L, H	2.80°	(parring)		and a second	31 pbs. of 4 variables	
A Pression	 Other 	Num.			a - 2	14 1441	(paring)		a ga. M	it phy, of 2 variables	
#Depression	4 1405.00	- Annual Co			W . U	Di sten i	terming		• m.eri	List of 14	
1.569	10 14	Male				a des	(and the second		• qs. ar34	List of II	
1.140		Male				LL date	terring.		In	200 obs. of 11 variables	
10 11 940	III Ibion	Dense	141			ta dan	(pawneg)		The second second second	(Marian)	
10 10 1000		New				0 - 10 V	(painting)				
12-140	-	Aless.				Ge then	Lasercing.				
10 11 140		- Herber				H-MA	(anoing)				
14 14 Late	in traini	And A	in the second		eff e	a day	queing		R Resources	RStudio	
name in that the owner									Learning & Online	AStacky IOE Support	
-										Estado Commany Fr	2411
Annual Article	ALC: NO.	CASH AND IN	No. of Concession,		COLUMN ST		Contrast of	Statement of the local division of the	R on MackOverfine	Stacky Own Sheets	
around editor years	a sector in	2010 104	Accession of						Getting Help with R	Ristuctio Tio of the Day	
									County in a main	Studio Packages	
THE PERSON NEWSFILM										Statio Products	
Citer - and Chine											
The surface of the su									Manuals		
									As introduction to R	The R Language Defini	20.1
weight and a									Writing II Columnian	E installation and Adm	
									R Data teport/Export	Ritenali	
Acat (Prov 103 Broup Scharts									Reference		
									Parkagen	Sharch Engine & Koyee	07028

###R Basics and Summary Statistics

1. Now that the data is loaded, we can evaluate it. First, we want to view some of the **summary statistics** for the dataset. Note, if at any time you want to write notes in your code to remind you what you're doing, you can do so by starting the notes with the # sign.



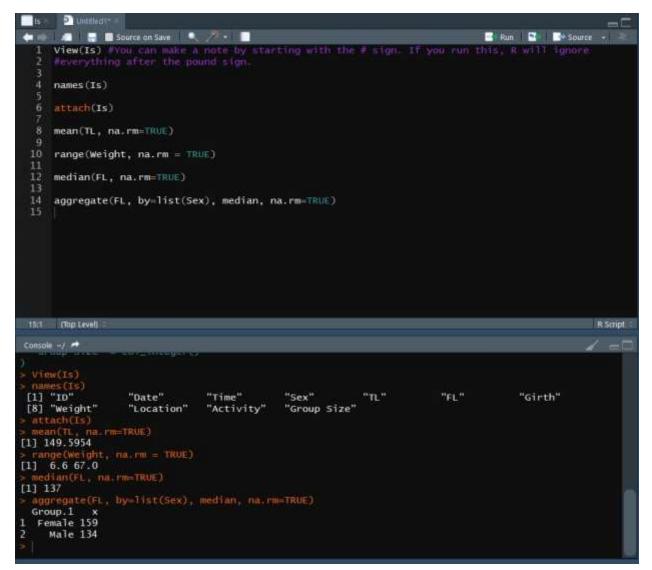
2. You can also tell R to accept, or "Run" your code by highlighting the code you want to use, then select the "Run" button top middle.

				Project	21100
Unificative P (unified)** • • Reservanyer, Prote * (Tetristation analysis R = • P Resson pair Res =	tree	ment	History		-
2 7 9 alon - 0 - Minut - 0	1	13 -	mport Dataiet + 🔮	11 Het	+1
36 - ###Importing Data	61	Giobal Ever	Q.		
22 Contraction of the second sec	Dat	8			
56 First, we need to import the data into m. 57	Q p	er, hats	h 12 obs. of 8 variables		10
Se just a couple of a basics before running code. If you use a pound sign (#) in code. It will make a		er, sort			12
convent. This convent is helpful for remembering what you did in the code. 39 To run 8 code, you highlight it and go to the upper right corner of this window and click "Run".			une. 07 mbs. of 6 variables		10
an and a second s second second se	1.0		umm. 72 obs. of 6 variables		10
41 m of 1		otal	96 obs. of 10 variables		
63 A It want to came or attract or enforce a same to re-	0.0	otairei	ated 72 obs. of 10 variables		. 10
44 Sense the data file (In Cher Set and Sense) 45 Sense (I) and the sense the data in the sense of the sense of the file of the sense. If yes right along the sense of the sense tables, result the sense of the backbor of the file.					
60 67 Ib- read, cov ("PATHWAY Adult 2018 data, cov")	1000	-	S		
48	The	1.00	Packages Help Viewer		-
69 70- ##WSummary Statistics	OL		0 Updata G		
21 「注意にはない」である「「加」になりためで、シントントロールの特徴になってないのないとなった。 パンパーロングの時代のないない	1000	Name	Departmen	Wens	
22 new that the data is loaded, we can play around with it, we're going to create some summary statistics for the data to see how it differs among groups.		paratlet	Support for Panelel computation in R	313	0
127 Q Devertant R 1 (R Markstown 1	T	rpart	Recursive Partitioning and	41-10	1.0
-D.	10	dentili-	Regression Trees		
of "license()" or "license()" for distribution details.	11	spatial	Functions for Ringing and Pairs Patrice Intelliges	73-11	. 9
		spinier	Regression Spline Functions and	33.3	0
in the company of the second se	-	and the	Classes		-
	닢	stats	The # Stat: Package	333	- 0
e 'contributors()' for more information and	불	Attak	Statistical Functions using 54 Classe		
se 'cuntributors()' for more information and itation()' on how to cite # or # packages in publications.		Survey a	San-weat Amatyses Tel/Th Interface	2.45-1	
pe 'constitutorSO' for more information and itation()' on how to cite # or # packages in publications. pe 'demu()' for mome demum, 'belp()' for an-line help, or elp.start() for an UNE demome 'inverface to belp.	12	hill.		333	. 0
is a collaborative project with many cumtributors. $p_{\rm constraints}(C)$ for seve information and itation() on how to cite a or a packages in publications. se down() for same down, help() for un-line help, or $p_{\rm p}$ start() for an effec broaser interface to belp. se $q()$ to quit a.	Ē		Tanh fan Factinge Oppstoprinet	333	- 0
se 'contributors()' for more information and itation()' on how to cite a or a packages in publications. se 'demu()' for some demus, 'belp()' for an-line help, or "by start()' for an UTM, demuser interfact to belp.		toob.			
se "centributors()" for more information and itation()" on how to cite m or m packages in publications. se "demu()" for mome demug, help()" for un-line help, or objectart()" for an WHG. browner interface to belp. m. "q()" to quit m.		toots translatio	The K Translations Package The K Uhly Package	111	

- 3. names(ls) Enter the names(ls) function into the coding window, then select "Run". The names() function tells R to retrieve the column names from your dataset. This is a good way to make sure that you have all the data you need. Also, when you're running code later, you can copy and paste the names from this output. R is case sensitive, so capitalizing when you didn't in your dataset can cause code to fail.
 - a. The names() function should output the following names in the **Console**: , "ID", "Date", "Time", "Sex", "TL", "FL", "Girth", "Weight", "Location", "Activity", "Group Size"
- 4. attach(ls) The attach() function allows us to tell R to consider ONLY the dataset we are working with at the current time. If you are working with multiple datasets, using the attach() function tells R to focus just on one. In the Coding Window, enter attach(ls) and select "Run"
- 5. Calculating the Mean Total Length (cm). Now that the data is attached, we can tell R to calculate the mean Total Length (cm) from all fish in 2018 using the mean() function. Sometimes, your data may be missing a measurement. It happens in the field. We can tell R to not consider cells in which there is no data using the na.rm = TRUE command. To calculate the mean Total Length, enter mean(TL, na.rm=TRUE) in the Coding Window, and select "Run". What is the mean TL for Lake Sturgeon in 2018? (149.5954 cm)
- Following the above format, we can determine the range in weights

 a. range(Weight, na.rm = TRUE) (6.6kg to 67.0kg)
- 7. And the median FL for male AND female lake sturgeon
 - a. median(FL, na.rm=TRUE) (137 cm)

- 8. If we want to evaluate summary statistics by groups, in our case by considering males and females as their own group, we can do that using the aggregate() function. Here, we're still telling R that we want to get a summary statistic, but we want to list that summary statistic by a certain grouping. In step 7, we calculated the median fork length for all sturgeon. Here, let's calculate the median fork length for male and female lake sturgeon. We are going to use the aggregate function to group by sex (by=list(Sex)), then insert the function we want (median)
 - a. Enter aggregate(FL, by=list(Sex), median, na.rm=TRUE) in the coding window, then **select "Run"**.
 - b. What are the median fork lengths for male and female lake sturgeon?
 - i. Females: 159 cm
 - ii. Males 134 cm



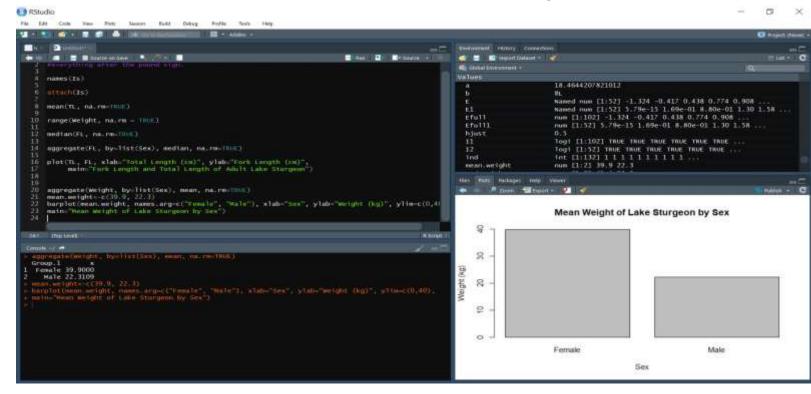
###Fast Visualizations

One of the most useful part of RStudio is the ability to see your data on a graph. Here, we will create some basic graphs which will allow us to see our data in order to help us detect differences. First, we will create a **scatter plot**. A **scatter plot** is a two dimensional graph which allows you to see each "fish" as the intersection of an x-axis variable and a y-axis variable in the form of a data point.

- 1. Create a scatterplot that shows FL and TL for each individual
 - a. We will use the plot() function to plot the Total Length (x-axis) of adult Lake Sturgeon against the Fork Length (y-axis). The plot function is written as follows for a **scatter plot**:
 - b. plot(x-axis variable name, y-axis variable name, "x-axis label", "y-axis label", "main graph title")
 - c. Since we are graphic Total Length vs Fork Length, our code should read:
 - i. plot(TL, FL, xlab="Total Length (cm)", ylab="Fork Length (cm)", main="Fork Length and Total Length of Adult Lake Sturgeon") **select "Run".**
 - 1. Where xlab= is the command which tells R the label of the x-axis;
 - 2. ylab= is the command which tells R the label of the y-axis;
 - 3. and main= is the commain which tells R the main chart label.
 - d. When you select "Run," you'll notice that a **scatter plot** appears in the bottom right box in Rstudio. You can download this graph, or simply view it in the window.

🖬 🐑 🍘 🖉 🛃 🦂 Anno 1970 - 1880 - Alba -				C Proved shown
A CONTRACTOR OF	49	Designer Mary Lies		Serie and the series of the se
• • A B Barrentas • / • B	a her a strategie			E Int + C
1. Vier(Is) who can have a mite by iterating with the P clubs If you re		🧉 📑 🖬 myzori Dalaint		Contraction of the second s
2 Prompting after the part then		Child Trevenue -	and the second se	A
		Femalenodel	List of 30	2
4 names(Is)		GLMMMigration	37 obs. of 11 variables Formal class glmerMod	
5 attach(ts)		a galish	List of 5	4
		ogs.ch23	List of 2	9
8 mean(TL, na.rm(T225)		egs.ch34	List of 2	20
10 range(veright, na.m = 1998)		Ogi-I	31 obs. of 4 variables	
		os.M	31 obs. of 2 variables	
12 sedian(FL, na.m=1000) 13		gs.mrl	List of 14	9
14 apprepate(FL, by-list(Sex), median, na.me-1000)		• gs.mr34	List of 11	0
		• 1s	200 obs. of 11 variables	
10 plot(TL, FL, klab-'Total Length (cs)", ylab-'Tork Length (cs)", 17 main-'Fork Length and Total Length of Adult Lake Stargnon")		A CONTRACTOR OF A CONTRACTOR O		
are starged and local Length or Maile Lake stargeon 3		The Mate Parkinger Ha		
118 /1				
		a a Phen Sh		
	in tanki in	For	k Length and Total Length of Adult	13 43
7.181 (Market)		For	k Length and Total Length of Adult	Lake Sturgeon
T Ten - Mig cavel - Canada	n bawi n	For	k Length and Total Length of Adult	Lake Sturgeon
The matched -	× -0	For	k Length and Total Length of Adult	Lake Sturgeon
Ten materit Constant - Vincitat - name(13) - name(13) - Tane", "The", "The", "The", "The", "The	× -0	For	k Length and Total Length of Adult	Lake Sturgeon
The Desteres Source of The Second Se	× -0	For	k Length and Total Length of Adult	Lake Sturgeon
The Obstant: State	× -0	For	k Length and Total Length of Adult	Lake Sturgeon
The Toolow' South a.e. Vinctic) * nont(Si) [1] '10" "Date" "Time" "Sex" "TL" "FL [8] 'Weight" "Location" "Activity" "Group Size" * stanf(Si) * sean(TL, instructNUE) [1] 140.954	× -0	For	k Length and Total Length of Adult	Lake Sturgeon
The Designer Constant View(c) const(c) (1) "10" "Date" "Time" "Sex" "TL" "TL (3) "Weight "Location" "Activity" "Group Size" * strain(C, ns, res=TRUE) D1 140.5954 * consecting(c, ns, res = TRUE)	× -0	For	k Length and Total Length of Adult	Lake Sturgeon
<pre> test matterst Crass view(it) name(it) "gate" "Time" "Sex" "TL" "TL (§] "Weight "Location" "Activity" "Group Size" stand(it) nameTRUE: [1] 140.595 second(it, na, rm = TRUE) independent(it, na, rm = TRUE) [1] 6.6 67.0 veediad(it, na, rm = TRUE) [1] 6.6 87.0 </pre>	× -0	For	k Length and Total Length of Adult	Lake Sturgeon
The Destrue: Sense =	× -0	Faik Length (cm)	k Length and Total Length of Adult	Lake Sturgeon
Test (matched) Constant of a View(it) resumm(15) (11 * 100 (11 * 100 * * ********************************	× -0	For	npart- 2	Lake Sturgeon
The matched : County Vice(it) remain(13) (11 * 100 * "The "Sex" "The "The" "Sex" "The ""Fh (8] 'Weight "Location" "Activity" "Group Size" * strain(15) * strai	× -0	Fork Length (cm) 100 120 140 160 180 100 120 140 160 180 	k Length and Total Length of Adult	Lake Sturgeon
Time Server	× -0	Faik Length (cm)	k Length and Total Length of Adult	Lake Sturgeon

- Another useful graph which allows you to compare two related groups of data to one another is a **bar graph or bar plot**. A **bar plot** allows you to compare a median or mean value of one group (ie: median FL of female sturgeon to male sturgeon) to another. Below, we are going to make a **bar plot** which does just that.
 - a. Make a **bar plot** of mean weight by Lake Sturgeon gender.
 - b. first we need to make an object that has the mean weight by each sex. We are going to use aggregate again
 - i. aggregate(Weight, by=list(Sex), mean, na.rm=TRUE) select "Run"
 - ii. females have a mean weight of 39.9 kg while males have a median weight of 22.3 kg.
 - iii. we're going to save these two values in an object called "mean.weight" using the following command:
 - 1. mean.weight<-c(39.9, 22.3) select "Run"
 - c. Now, we are going to use the barplot() function to create the graph. We are graphing the mean.weight object we just made.
 - i. barplot(mean.weight, names.arg=c("Female", "Male"), xlab="Sex", ylab="Weight (kg)", ylim=c(0,40), main="Mean Weight of Lake Sturgeon by Sex")
 select "Run"
 - d. Where:
 - i. Names.arg is the labels that go under the column values;
 - ii. xlab= and ylab= are the x and y axes lables;
 - iii. ylim= sets the scale of the y-axis from 0 to 40; and
 - iv. main= is the title of the graph
 - e. Again, the **bar plot** should appear in the lower right box. It appears that there's a pretty noticeable difference between male and female median weight.



###Correlation

Graphs are very useful in detecting difference in data, but how do we know that those relationships can be quantified? In the first example, we looked at the visual relationship between fork length and total length. They look like their related, in when total length increases on the x-axis, fork length increases on the y-axis. But how do we know if that's ACTUALLY what's going on?

We can use a simple **correlation analysis** to determine if two variables are related. A **correlation** analysis simply tells us whether two variables track one another. We have to be careful with correlation. If we consider the variables with which we work, it's great. But we should remember that correlation does NOT equal causation. In other words, we can have a scatter plot which tells us that two variables appear to be related, when in fact, they are obviously not (Figure 4). As long as you are careful when

considering the relationship two variables have, correlation can be a powerful investigative tool. A **correlation analysis** outputs a single value between 0 and 1 called the **correlation coefficient**. The closer to 1 a **correlation coefficient** falls; the more correlated two variables are. Conversely, as a correlation coefficient approaches 0, the more poorly correlated variables are to one another.

Here, we will consider how variables from the **scatter plot** example are actually related.

- 1. Find the correlation factor between TL and FL.
 - Use the cor() function for correlation. We are correlating FL and TL. In the case where we may

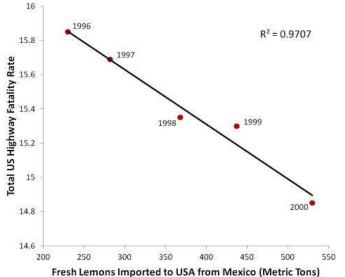


Figure 4. Correlation ≠ causation. Is there a reasonable relationship between the number of lemons imported from Mexico and the number of US Highway Fatalities....?

be missing data from either total length, or fork length we can use the command use="complete.obs" to tell R to ignore incomplete data.

- b. In the Coding Window, enter: cor(FL, TL, use="complete.obs") select "Run"
- c. The **correlation coefficient** is 0.9894 which is very close to 1, so these two variables are highly correlated. This makes sense as FL is a very similar measurement to TL.

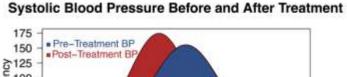
###T-Test

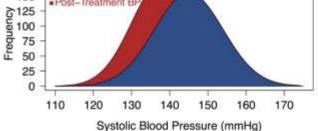
In the second example, we considered the average weight of male and female Lake Sturgeon. It appeared that females weighed more than males. However, how do we know that? What if the majority of your female population wasn't all that different from the male population, but you happened to catch three female Lake Sturgeon that were **outliers**? An **outlier** is a data point which is distant or well removed from other observations. We need a way to determine if the data we consider are **statistically different.** In other words, can we use basic statistics to determine if two groups are quantifiably

different, or just visually different. In the case of our weight example, we can use a **T-Test.** A **T-Test** compares the mean values of a variables between two groups. It does that by calculating a t-statistic and a p-value. Now, rather than explain the math, let's instead consider a visual explanation. Figure 5

shows distributions of blood pressure measurements prior to and after a treatment. We do that by interpreting a **p-value**. For the sake of this lesson, a **pvalue** less than 0.05 (P < 0.05) would tell us that our considered groups are statistically different.

> Use a t-test to statistically compare the mean weights of male and female lake sturgeon caught in 2018.





- 2. Use the t.test() function to see if Weight differs by (~) sex
- 3. Enter: t.test(Weight ~ Sex) select "Run"
- 4. We see a lot of information in the output from the t-test code (see below). The important values are in boxes. As above, we can see that the average weight of female Lake Sturgeon is 39.9 kg, which the average weight of males is 22.3 kg. Additionally, we can see in the green box that the calculated p-value in 8.48 x 10⁻¹³. This value is far less than the threshold of 0.05 mentioned above. As such, we can interpret that the mean weight of female Lake Sturgeon is different than the mean weight of male Lake Sturgeon. Our graph shows us that female Lake Sturgeon weight more than male Lake Sturgeon.

26 27 t.test(weight ~ 28	Sex)	
¢		
28:1 (Top Level) 1		R Script
Console -/ 🔿		
> t.test(Weight ~ Sex))(c)	
Welch Two Samp	le t-test	
data: Weight by Sex t = 9.3774, df = 52.39 alternative hypothesis 95 percent confidence 13.82591 21.35230 sample estimates: mean in group Female	2, <u>p-value = 8.481e-13</u> : true difference in means is not equal to 0 interval: mean in group Male	
39.9000	22.3109	
*		

###Assignment

1. What is the minimum and maxium (range) total length (cm) of male and female lake sturgeon captured in 2018?

Teacher Answer:

Code (Both combined): range(TL, na.rm = TRUE)

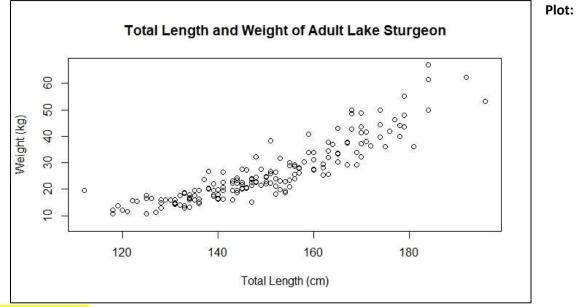
Answer: min – 112cm; max 196 cm.

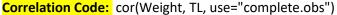
Code (Seperated by group): aggregate(TL, by=list(Sex), range, na.rm=TRUE)

Answer: Females: 138cm – 196cm; Males: 112cm – 178 cm

2. Create a scatter plot and correlation coefficient for Total Length (x-axis) and Weight (y-axis) of adult lake sturgeon captured in 2018. Are Total Length and Weight correlated?

Graph Code: plot(TL, Weight, xlab="Total Length (cm)", ylab="Weight (kg)", main="Total Length and Weight of Adult Lake Sturgeon")



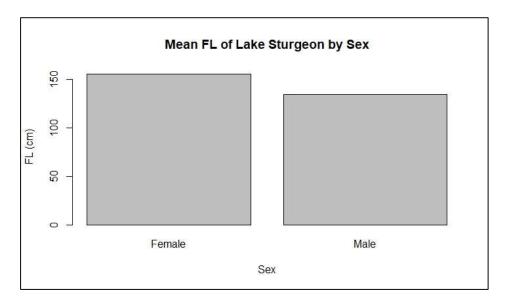


Teacher Answer: Correlation coefficient: 0.8947 – yes they are correlated.

3. Calculate the mean fork length for male and female Lake Sturgeon. Create a barplot showing the means of the two groups. Use a T-Test to determine if fork length differs between male and female Lake Sturgeon.

Mean Code:

aggregate(FL, by=list(Sex), mean, na.rm=TRUE) mean.FL<-c(155.64, 134.23) Graph Code: barplot(mean.FL, names.arg=c("Female", "Male"), xlab="Sex", ylab="FL (cm)", ylim=c(0,160), main="Mean FL of Lake Sturgeon by Sex")



T-Test Code: t.test(FL ~ Sex)

mean in group Female mean in group Male

155.6389 134.2263

Teacher Answer: p-value = 4.10 x 10⁻¹¹; because p<0.05, we can determine that Fork Length differs between male and female Lake Sturgeon

4. Come up with a simple question you could answer with the 2018 adult capture data. This question must be different than the analyses performed above. Answer your question using at least 2 of the following: descriptive statistics, visualizations, statistical analyses.

Teacher Answer: Answers will vary. Could consider numerous relationships via correlation/scatter plot or barplot/T-test. As long as the code is similar to above, should be fine.