Tip!

Before you begin:
Print out a hard copy of this document to use as you follow along.

Describing the Relationship between TWO Variables

Introduction:

This is a "show me" illustration for BIOSTATS 540 ***Unit 1. Summarizing Data*** that utilizes the online tool **StatKey** which is found on www.lock5stat.com. This "show me" gives you experience **describing the relationship between two variables** in two settings:

(1) One variable **categorical**, one variable quantitative **continuous**, and
(4) **Both** variables quantitative **continuous**.

Data

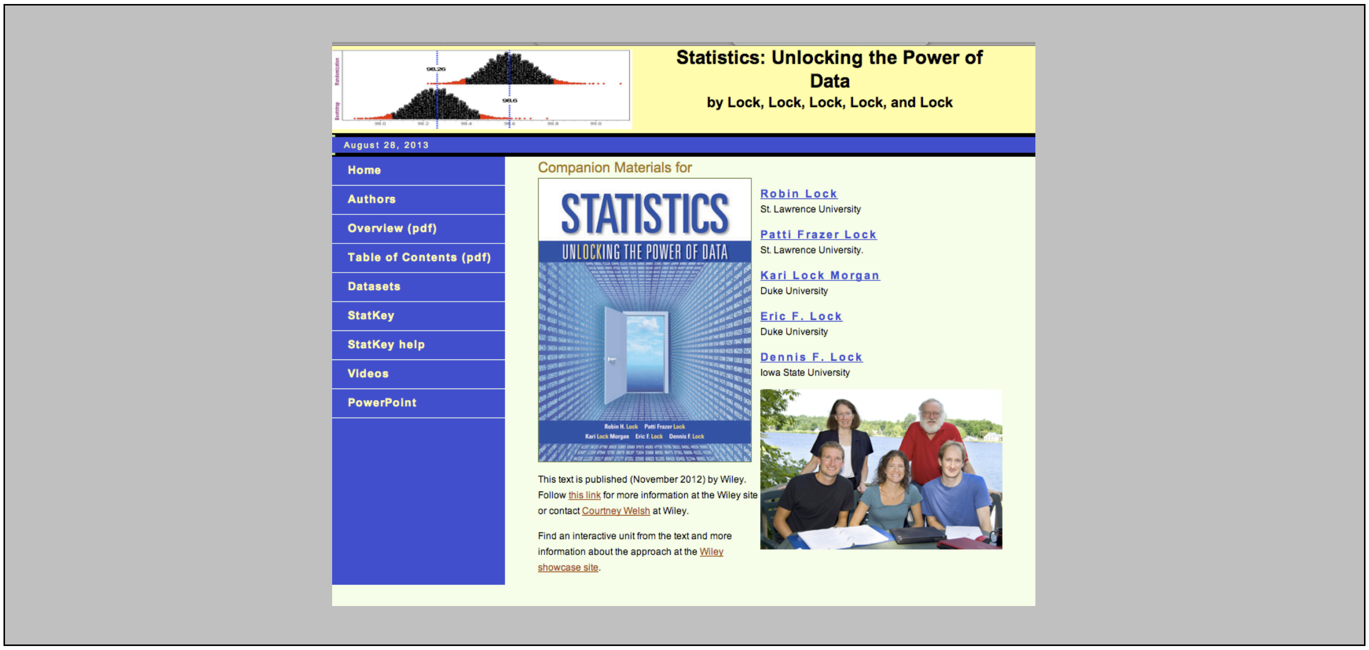
Source:
Lock RH, Lock PF, Morgan KL, Lock EF and Lock DF. *Statistics: Unlocking the Power of Data* New York, John Wiley, 2013.
Note – it is NOT necessary to purchase this book.

The data set used is called **NutritionStudy** and is freely available for download (in multiple formats) from the Lock, Lock, Morgan, Lock and Lock book. We will be working with the excel version, **Nutritionstudy.xls.**

NutritionStudy.xls contains information on 315 individuals. Apart from study id, for each individual there are 16 variable measurements. Some are dietary. Others are concentrations of selected micronutrients in the blood.

Activity #1. Launch lock5stat and download **NutritionStudy.xls** to your computer.

__1. Using whatever browser you have, launch http://www.lock5stat.com. A home page will appear:



Statistics: Unlocking the Power of Data

by Lock, Lock, Lock, Lock, and Lock

August 28, 2013

Companion Materials for

Robin Lock
St. Lawrence University

Patti Frazer Lock
St. Lawrence University.

Kari Lock Morgan
Duke University

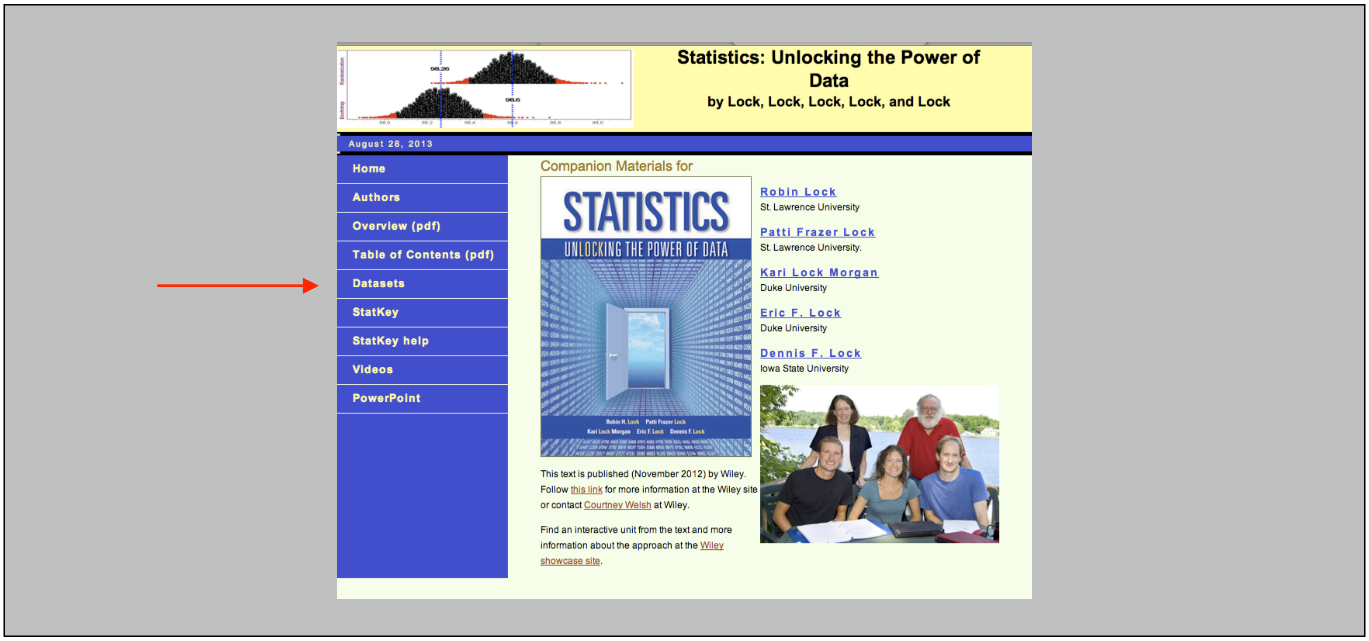
Eric F. Lock
Duke University

Dennis F. Lock
Iowa State University

This text is published (November 2012) by Wiley. Follow this link for more information at the Wiley site or contact Courtney Welsh at Wiley.

Find an interactive unit from the text and more information about the approach at the Wiley showcase site.

__1. From the home page of http://www.lock5stat, click at left on the button, **Datasets**.



Statistics: Unlocking the Power of Data

by Lock, Lock, Lock, Lock, and Lock

August 28, 2013

Companion Materials for

Robin Lock
St. Lawrence University

Patti Frazer Lock
St. Lawrence University.

Kari Lock Morgan
Duke University

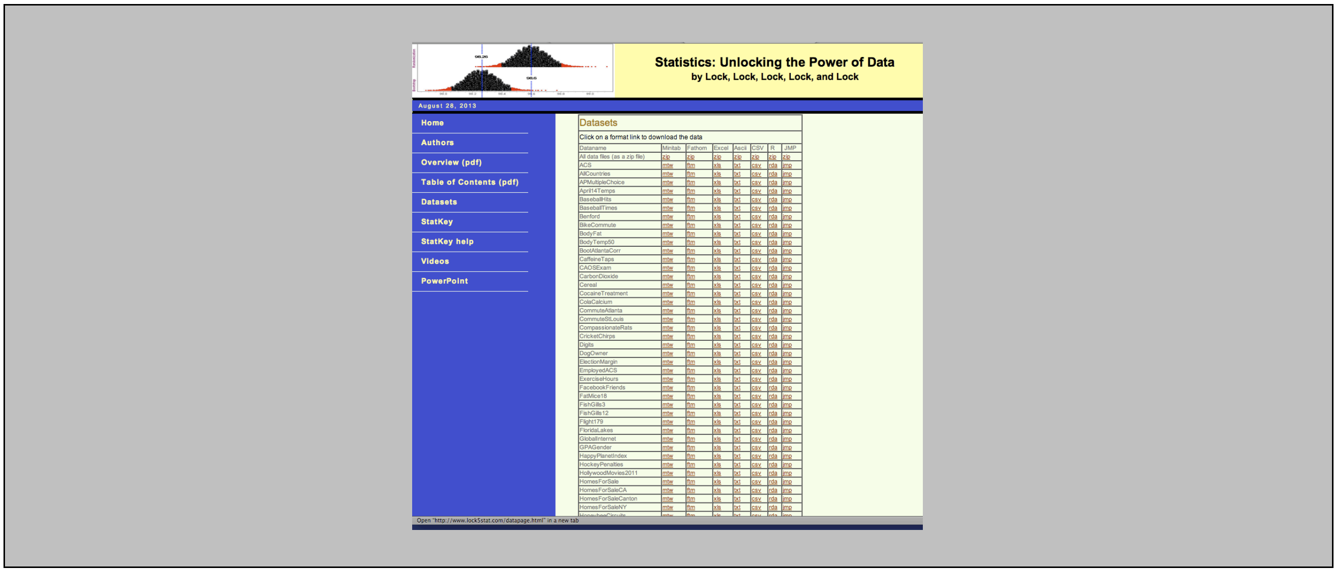
Eric F. Lock
Duke University

Dennis F. Lock
Iowa State University

This text is published (November 2012) by Wiley. Follow this link for more information at the Wiley site or contact Courtney Welsh at Wiley.

Find an interactive unit from the text and more information about the approach at the Wiley showcase site.

You will see a long table of data sets. They are listed in alphabetic order by dataset name.



Statistics: Unlocking the Power of Data
by Lock, Lock, Lock, Lock, and Lock

August 28, 2013

Datasets

Click on a format link to download the data

Dataname	Minitab	Fathom	Excel	Ascii	CSV	R	JMP
All data files (as a zip file)	zip	zip	zip	zip	zip	zip	zip
ACS	mtw	ftm	xls	txt	csv	rda	jmp
AllCountries	mtw	ftm	xls	txt	csv	rda	jmp
APMultipleChoice	mtw	ftm	xls	txt	csv	rda	jmp
April14Temps	mtw	ftm	xls	txt	csv	rda	jmp
BaseballHits	mtw	ftm	xls	txt	csv	rda	jmp
BaseballTimes	mtw	ftm	xls	txt	csv	rda	jmp
Benford	mtw	ftm	xls	txt	csv	rda	jmp
BikeCommute	mtw	ftm	xls	txt	csv	rda	jmp
BodyFat	mtw	ftm	xls	txt	csv	rda	jmp
BodyTemp50	mtw	ftm	xls	txt	csv	rda	jmp
BootAtlantaCorr	mtw	ftm	xls	txt	csv	rda	jmp
CaffeineTaps	mtw	ftm	xls	txt	csv	rda	jmp
CAOSExam	mtw	ftm	xls	txt	csv	rda	jmp
CarbonDioxide	mtw	ftm	xls	txt	csv	rda	jmp
Cereal	mtw	ftm	xls	txt	csv	rda	jmp
CocaineTreatment	mtw	ftm	xls	txt	csv	rda	jmp
ColaCalcium	mtw	ftm	xls	txt	csv	rda	jmp
CommuteAtlanta	mtw	ftm	xls	txt	csv	rda	jmp
CommuteStLouis	mtw	ftm	xls	txt	csv	rda	jmp
CompassionateRats	mtw	ftm	xls	txt	csv	rda	jmp
CricketChirps	mtw	ftm	xls	txt	csv	rda	jmp
Digits	mtw	ftm	xls	txt	csv	rda	jmp
DogOwner	mtw	ftm	xls	txt	csv	rda	jmp
ElectionMargin	mtw	ftm	xls	txt	csv	rda	jmp
EmployedACS	mtw	ftm	xls	txt	csv	rda	jmp
ExerciseHours	mtw	ftm	xls	txt	csv	rda	jmp
FacebookFriends	mtw	ftm	xls	txt	csv	rda	jmp
FatMice18	mtw	ftm	xls	txt	csv	rda	jmp
FishGills3	mtw	ftm	xls	txt	csv	rda	jmp
FishGills12	mtw	ftm	xls	txt	csv	rda	jmp
Flight179	mtw	ftm	xls	txt	csv	rda	jmp
FloridaLakes	mtw	ftm	xls	txt	csv	rda	jmp
GlobalInternet	mtw	ftm	xls	txt	csv	rda	jmp
GPAGender	mtw	ftm	xls	txt	csv	rda	jmp
HappyPlanetIndex	mtw	ftm	xls	txt	csv	rda	jmp
HockeyPenalties	mtw	ftm	xls	txt	csv	rda	jmp
HollywoodMovies2011	mtw	ftm	xls	txt	csv	rda	jmp
HomesForSale	mtw	ftm	xls	txt	csv	rda	jmp
HomesForSaleCA	mtw	ftm	xls	txt	csv	rda	jmp
HomesForSaleCanton	mtw	ftm	xls	txt	csv	rda	jmp
HomesForSaleNY	mtw	ftm	xls	txt	csv	rda	jmp
HoneybeeCircuits	mtw	ftm	xls	txt	csv	rda	jmp

Open "http://www.lock5stat.com/datapage.html" in a new tab

__3. Scroll down to locate the row for **NutritionStudy**.
To download this data to your computer, **click on the entry** with the extension **.xls**.

__4. Your browser will provide some sort of drop down menu, select Download File. **Note – Don't panic, here. The options in your drop down menu might be slightly different, depending on whether you are a PC user or a MAC user. Play with the possibilities here.**

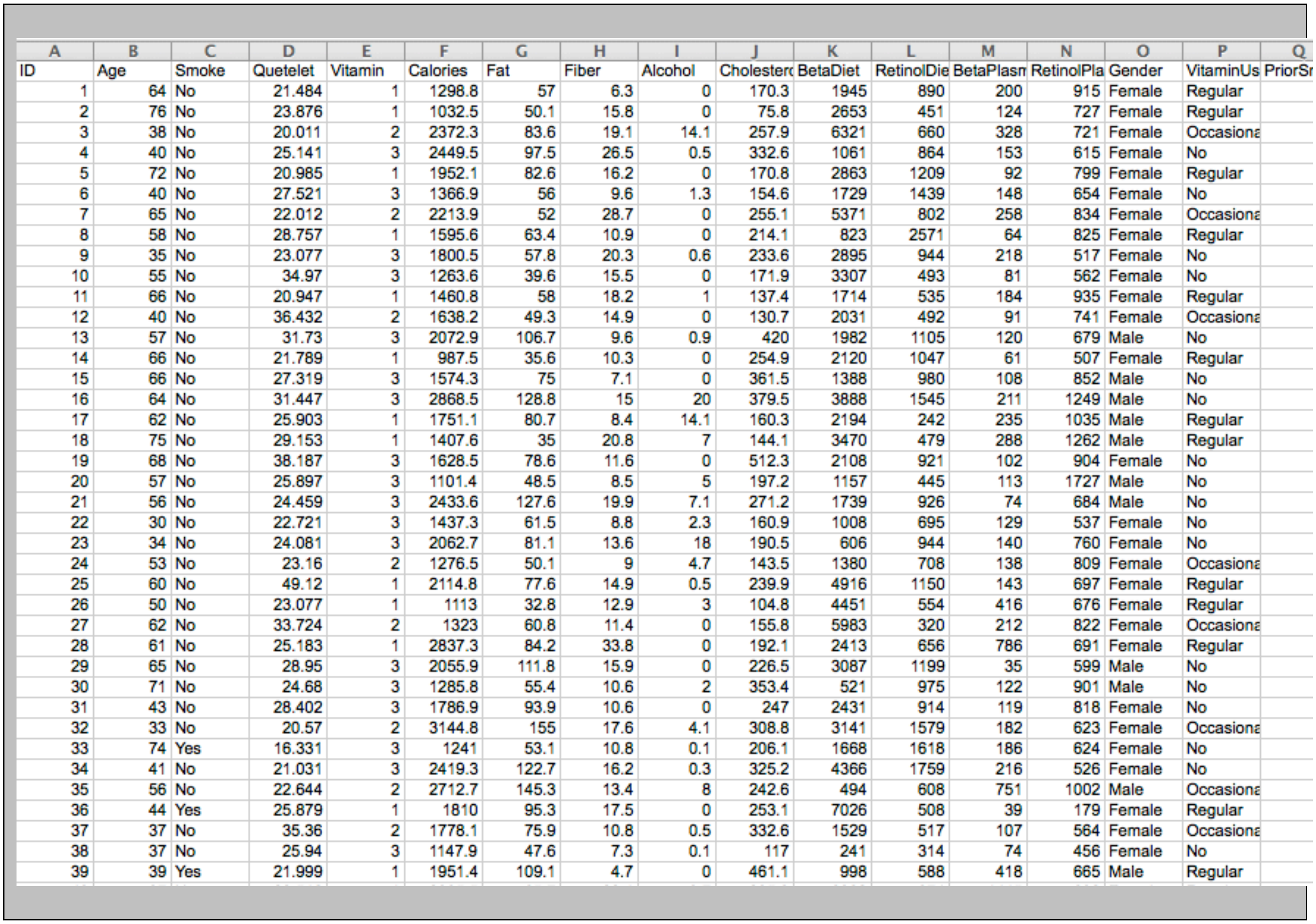
__5. Download **Nutrition.xls** to your desktop, or to some other location, that is easy for you to remember.

Following is a data dictionary (coding manual) for these data.

Variable Name	Label	Coding/Remarks
ID	ID for each subject in sample	Coded 1, 2, etc
Age	Subjects age, in whole years	Example: 64
Smoke	Does the subject smoke	Yes, No
Quetelet	Weight/(Height^2)	Example: 21.4838
Vitamin	Vitamin Use	1=Regular, 2=Occasional, 3=No
Calories	Number of calories consumed per day	Example: 1298.8
Fat	Grams of fat consumed per day	Example: 50.1
Fiber	Grams of fiber consumed per day	Example: 15.8
Alcohol	Average # of drinks consumed per week	Example: 1.3
Cholesterol	Cholesterol consumed per day (mg)	Example: 170.3
BetaDiet	Beta-carotene consumed per day (mcg)	Example: 1945
BetaPlasma	Plasma Concentration beta-carotene (ng/ml)	Example: 328
RetinolPlasma	Plasma Concentration retinon (ng/ml)	Example: 721
Gender	Female or Male	Female, Male
VitaminUse	Vitamin use	Coded as No, Occasional, Regular
PriorSmoke	Smoking Status	1=Never, 2=Former, 3=Current

Activity #2. In a separate window that you will keep open, launch excel and open the file NutitionStudy.xls. Create 2 new spreadsheets within this excel file workbook, one for each scenario.

__1. You should see the following. **Note – Yours might not look exactly the same, depending on whether you are a PC user or a MAC user and depending on which version of EXCEL you are using.**



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
ID	Age	Smoke	Quetelet	Vitamin	Calories	Fat	Fiber	Alcohol	Cholesterc	BetaDiet	RetinolDie	BetaPlasn	RetinolPla	Gender	VitaminUs	PriorSr
1	64	No	21.484	1	1298.8	57	6.3	0	170.3	1945	890	200	915	Female	Regular	
2	76	No	23.876	1	1032.5	50.1	15.8	0	75.8	2653	451	124	727	Female	Regular	
3	38	No	20.011	2	2372.3	83.6	19.1	14.1	257.9	6321	660	328	721	Female	Occasiona	
4	40	No	25.141	3	2449.5	97.5	26.5	0.5	332.6	1061	864	153	615	Female	No	
5	72	No	20.985	1	1952.1	82.6	16.2	0	170.8	2863	1209	92	799	Female	Regular	
6	40	No	27.521	3	1366.9	56	9.6	1.3	154.6	1729	1439	148	654	Female	No	
7	65	No	22.012	2	2213.9	52	28.7	0	255.1	5371	802	258	834	Female	Occasiona	
8	58	No	28.757	1	1595.6	63.4	10.9	0	214.1	823	2571	64	825	Female	Regular	
9	35	No	23.077	3	1800.5	57.8	20.3	0.6	233.6	2895	944	218	517	Female	No	
10	55	No	34.97	3	1263.6	39.6	15.5	0	171.9	3307	493	81	562	Female	No	
11	66	No	20.947	1	1460.8	58	18.2	1	137.4	1714	535	184	935	Female	Regular	
12	40	No	36.432	2	1638.2	49.3	14.9	0	130.7	2031	492	91	741	Female	Occasiona	
13	57	No	31.73	3	2072.9	106.7	9.6	0.9	420	1982	1105	120	679	Male	No	
14	66	No	21.789	1	987.5	35.6	10.3	0	254.9	2120	1047	61	507	Female	Regular	
15	66	No	27.319	3	1574.3	75	7.1	0	361.5	1388	980	108	852	Male	No	
16	64	No	31.447	3	2868.5	128.8	15	20	379.5	3888	1545	211	1249	Male	No	
17	62	No	25.903	1	1751.1	80.7	8.4	14.1	160.3	2194	242	235	1035	Male	Regular	
18	75	No	29.153	1	1407.6	35	20.8	7	144.1	3470	479	288	1262	Male	Regular	
19	68	No	38.187	3	1628.5	78.6	11.6	0	512.3	2108	921	102	904	Female	No	
20	57	No	25.897	3	1101.4	48.5	8.5	5	197.2	1157	445	113	1727	Male	No	
21	56	No	24.459	3	2433.6	127.6	19.9	7.1	271.2	1739	926	74	684	Male	No	
22	30	No	22.721	3	1437.3	61.5	8.8	2.3	160.9	1008	695	129	537	Female	No	
23	34	No	24.081	3	2062.7	81.1	13.6	18	190.5	606	944	140	760	Female	No	
24	53	No	23.16	2	1276.5	50.1	9	4.7	143.5	1380	708	138	809	Female	Occasiona	
25	60	No	49.12	1	2114.8	77.6	14.9	0.5	239.9	4916	1150	143	697	Female	Regular	
26	50	No	23.077	1	1113	32.8	12.9	3	104.8	4451	554	416	676	Female	Regular	
27	62	No	33.724	2	1323	60.8	11.4	0	155.8	5983	320	212	822	Female	Occasiona	
28	61	No	25.183	1	2837.3	84.2	33.8	0	192.1	2413	656	786	691	Female	Regular	
29	65	No	28.95	3	2055.9	111.8	15.9	0	226.5	3087	1199	35	599	Male	No	
30	71	No	24.68	3	1285.8	55.4	10.6	2	353.4	521	975	122	901	Male	No	
31	43	No	28.402	3	1786.9	93.9	10.6	0	247	2431	914	119	818	Female	No	
32	33	No	20.57	2	3144.8	155	17.6	4.1	308.8	3141	1579	182	623	Female	Occasiona	
33	74	Yes	16.331	3	1241	53.1	10.8	0.1	206.1	1668	1618	186	624	Female	No	
34	41	No	21.031	3	2419.3	122.7	16.2	0.3	325.2	4366	1759	216	526	Female	No	
35	56	No	22.644	2	2712.7	145.3	13.4	8	242.6	494	608	751	1002	Male	Occasiona	
36	44	Yes	25.879	1	1810	95.3	17.5	0	253.1	7026	508	39	179	Female	Regular	
37	37	No	35.36	2	1778.1	75.9	10.8	0.5	332.6	1529	517	107	564	Female	Occasiona	
38	37	No	25.94	3	1147.9	47.6	7.3	0.1	117	241	314	74	456	Female	No	
39	39	Yes	21.999	1	1951.4	109.1	4.7	0	461.1	998	588	418	665	Male	Regular	

__2. **FIRST SPREADSHEET** - In Excel, create a new sheet called **smoke and quetelet**:

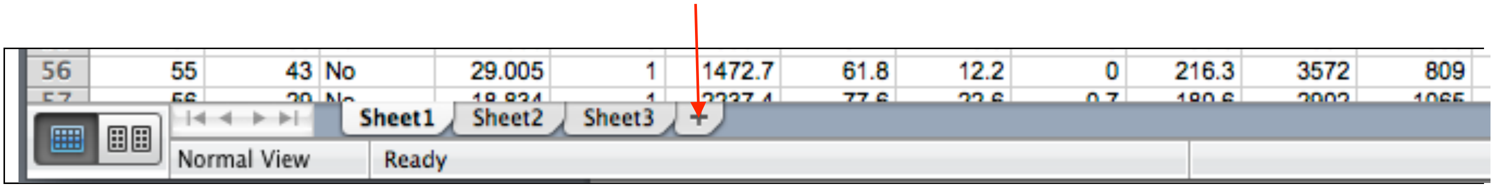
__ Preliminary – Notice that you are located in Sheet1

__ While in Sheet 1, position your cursor over the "C" at the top and that refers to column "C" (Smoke).

__ (Copy) Select this entire column of data by clicking on the "C" that is the column heading

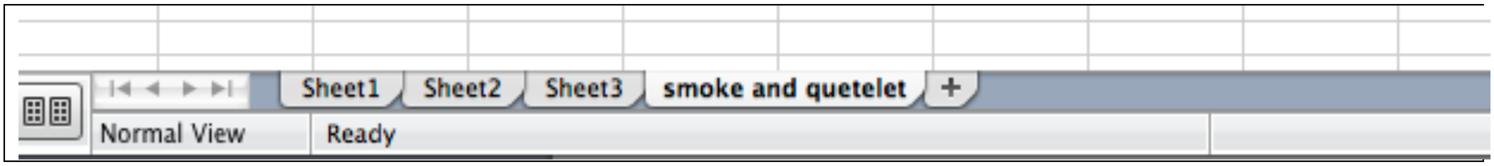
__ (Copy) From the Excel main menu, **EDIT > COPY**

__ Next, at the bottom of your screen, click on the "+". This will bring you to a new sheet



__ Preliminary – Notice that you are now located in Sheet4

__ While in Sheet4, at bottom, right click on Sheet4. Select rename. Rename **smoke and quetelet**



__ While in Sheet **smoke and quetelet**, position your cursor in cell Row:1 Column: A.

__ (Paste) From the Excel main menu, **EDIT > PASTE**

Now obtain the data on quetelet (note – possibly you copied 2 columns at once already):

__ Activate **Sheet1**, position your cursor over the "D" at the top and that refers to column "D" (Quetelet).

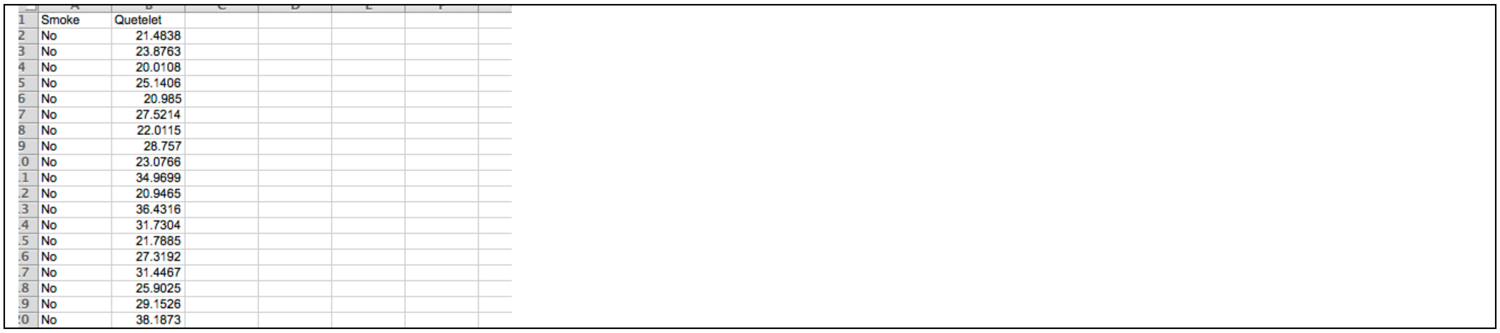
__ (Copy) Select this entire column of data by clicking on the "D" that is the column heading

__ (Copy) From top menu, EDIT > COPY

__ Activate sheet **smoke and quetelet**, position your cursor in Row:1 Column: B.

__ (Paste) From top menu, EDIT > PASTE.

You should now see:

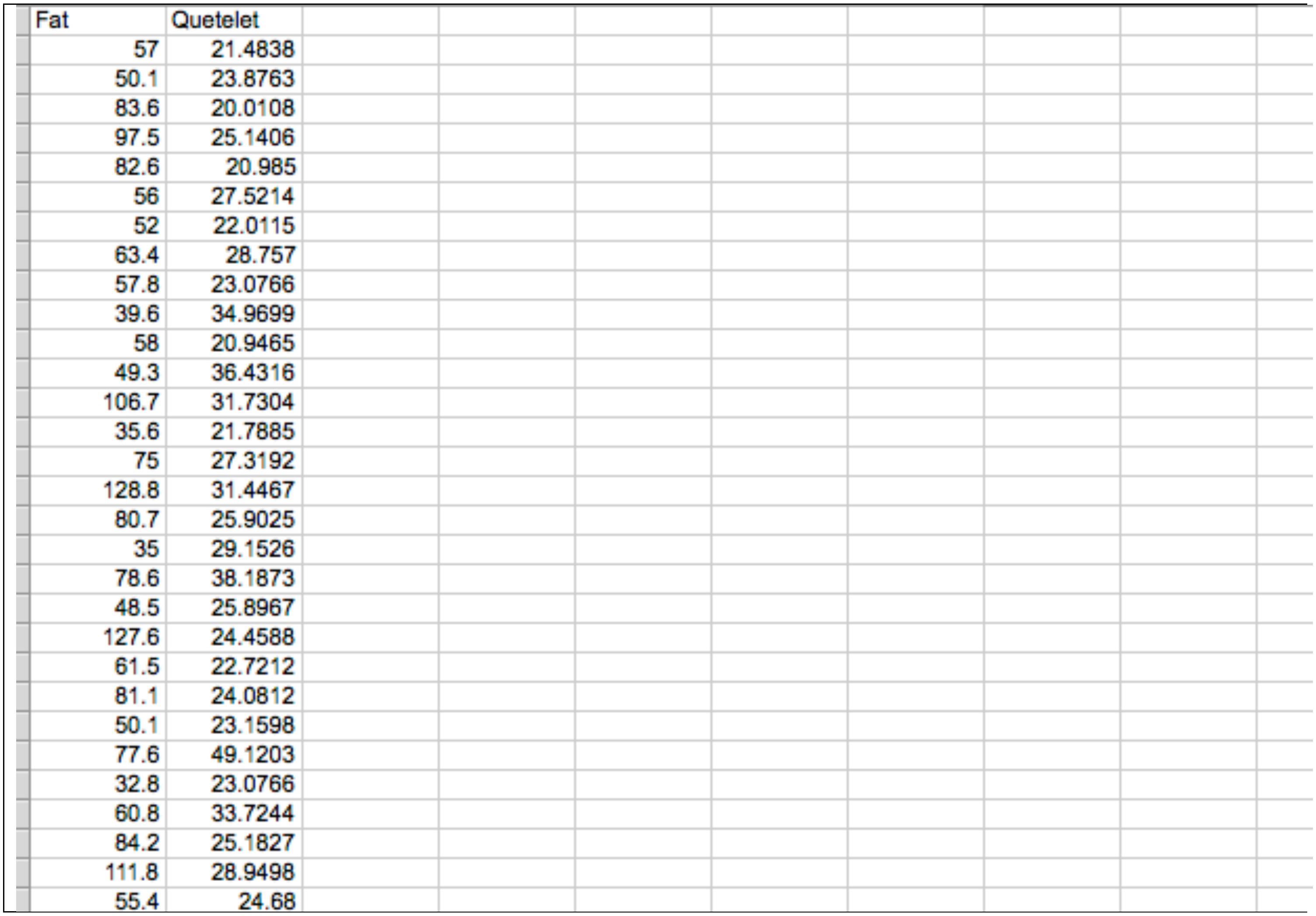


1	Smoke	Quetelet
2	No	21.4838
3	No	23.8763
4	No	20.0108
5	No	25.1406
6	No	20.985
7	No	27.5214
8	No	22.0115
9	No	28.757
0	No	23.0766
1	No	34.9699
2	No	20.9465
3	No	36.4316
4	No	31.7304
5	No	21.7885
6	No	27.3192
7	No	31.4467
8	No	25.9025
9	No	29.1526
0	No	38.1873

Save your work: In Excel, from main menu: **FILE > SAVE AS**

__3. **SECOND SPREADSHEET**
In Excel, following the same steps as detailed in step 2, create a new sheet called **fat and quetelet**:
Fat data is in column **G**
Quetelet data is in column **D**

You should now see:



Fat	Quetelet
57	21.4838
50.1	23.8763
83.6	20.0108
97.5	25.1406
82.6	20.985
56	27.5214
52	22.0115
63.4	28.757
57.8	23.0766
39.6	34.9699
58	20.9465
49.3	36.4316
106.7	31.7304
35.6	21.7885
75	27.3192
128.8	31.4467
80.7	25.9025
35	29.1526
78.6	38.1873
48.5	25.8967
127.6	24.4588
61.5	22.7212
81.1	24.0812
50.1	23.1598
77.6	49.1203
32.8	23.0766
60.8	33.7244
84.2	25.1827
111.8	28.9498
55.4	24.68

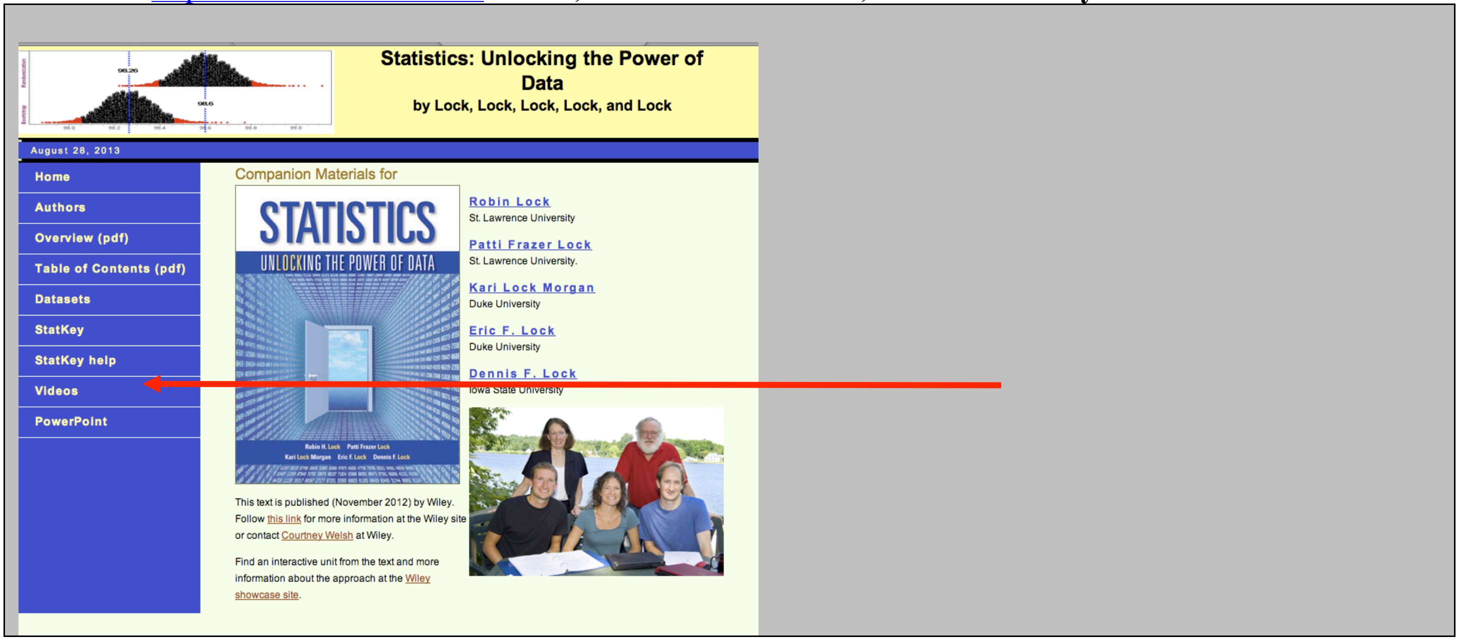
Again, save your work: In Excel, from top menu bar: **FILE > SAVE AS**

__4. **Minimize Excel** window
__ but ***do NOT exit*** Excel. We will be coming back to this.

Summarizing a Relationship Between Two Variables
One Categorical (smoke), One Quantitative Continuous (quetelet)

Activity #3. Launch the StatKey tool. Then, under *Descriptive Statistics and Graphs*, click *One Quantitative and One Categorical Variable*

__1. Re-launch http://www.lock5stat.com. Then, from the menu at left, click on **StatKey:**



Statistics: Unlocking the Power of Data
by Lock, Lock, Lock, Lock, and Lock

August 28, 2013

Companion Materials for

Robin Lock
St. Lawrence University

Patti Frazer Lock
St. Lawrence University.

Kari Lock Morgan
Duke University

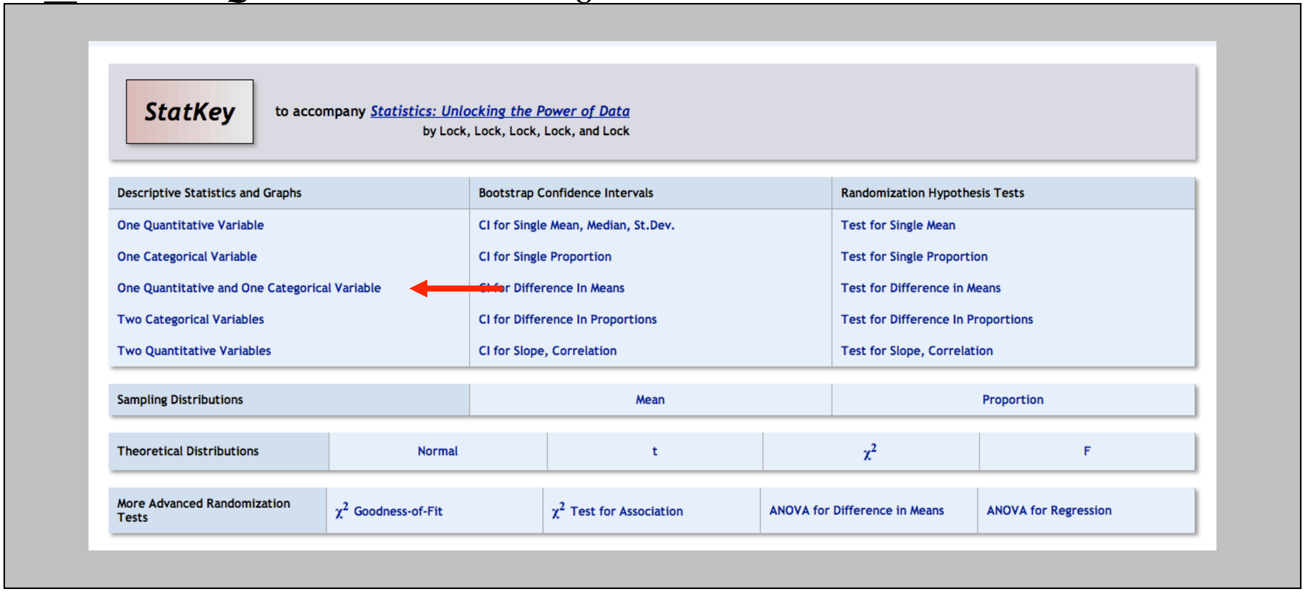
Eric F. Lock
Duke University

Dennis F. Lock
Iowa State University

This text is published (November 2012) by Wiley. Follow this link for more information at the Wiley site or contact Courtney Welsh at Wiley.

Find an interactive unit from the text and more information about the approach at the Wiley showcase site.

___2. From the selection of analysis options offered, under *Descriptive Statistics and Graphs*,
__ click ***One Quantitative and One Categorical Variable***



StatKey to accompany *Statistics: Unlocking the Power of Data*
by Lock, Lock, Lock, Lock, and Lock

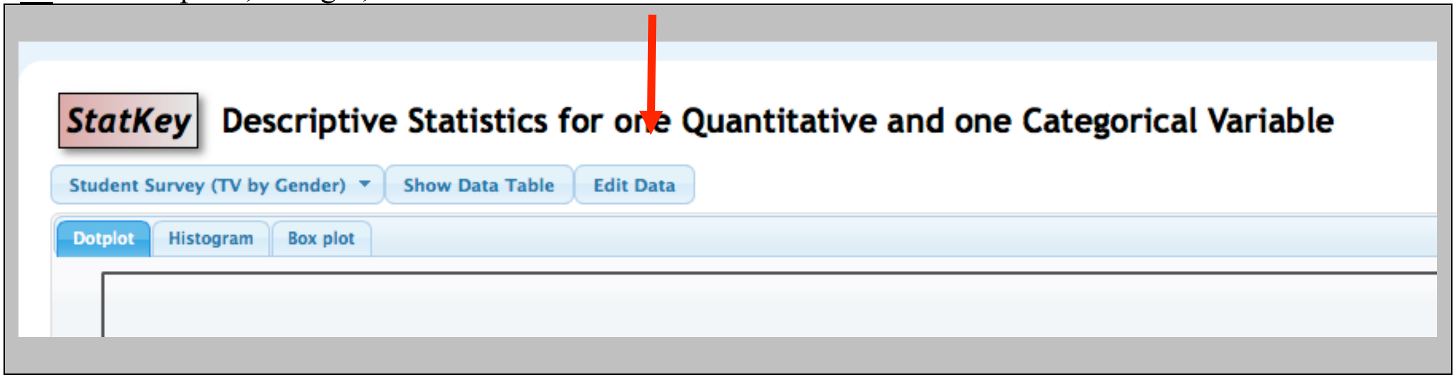
Descriptive Statistics and Graphs	Bootstrap Confidence Intervals	Randomization Hypothesis Tests
One Quantitative Variable	CI for Single Mean, Median, St.Dev.	Test for Single Mean
One Categorical Variable	CI for Single Proportion	Test for Single Proportion
One Quantitative and One Categorical Variable	CI for Difference In Means	Test for Difference in Means
Two Categorical Variables	CI for Difference In Proportions	Test for Difference In Proportions
Two Quantitative Variables	CI for Slope, Correlation	Test for Slope, Correlation

Sampling Distributions	Mean	Proportion

Theoretical Distributions	Normal	t	χ^2	F

More Advanced Randomization Tests	χ^2 Goodness-of-Fit	χ^2 Test for Association	ANOVA for Difference in Means	ANOVA for Regression

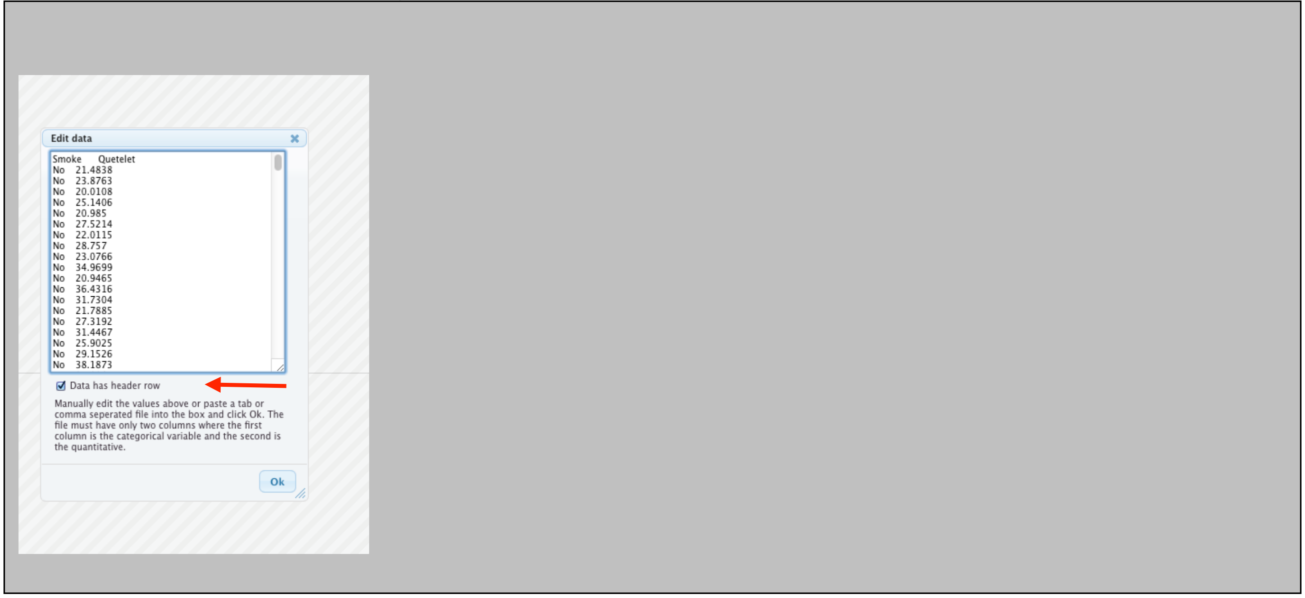
___3. Do the following to replace the default data with the data on smoke and quetelet.
__ At the top bar, on right, click **Edit data**


StatKey
Descriptive Statistics for one Quantitative and one Categorical Variable
Student Survey (TV by Gender)
Show Data Table
Edit Data
Dotplot
Histogram
Box plot

___4. ***In StatKey***, delete the default data by doing the following.
___ Using your cursor, position and drag to select all the default.
___ Click on the delete key on your keyboard.
___ ***Important*** – Do ***NOT*** click the ok button just yet.

___5. ***In Excel***, select the smoke and quetelet data .
___ Using your cursor, position and drag to select all the default.
___ From main menu at top, EDIT > COPY.
___ ***Important*** – Minimize Excel, Do ***NOT*** exit.

___6. ***In StatKey,*** from main menu of your browser use EDIT > PASTE to paste in the data values
__ BOX CHECK: Check to see that the box next to "Data has a header row" is checked.
__ Click on **OK** at bottom right.

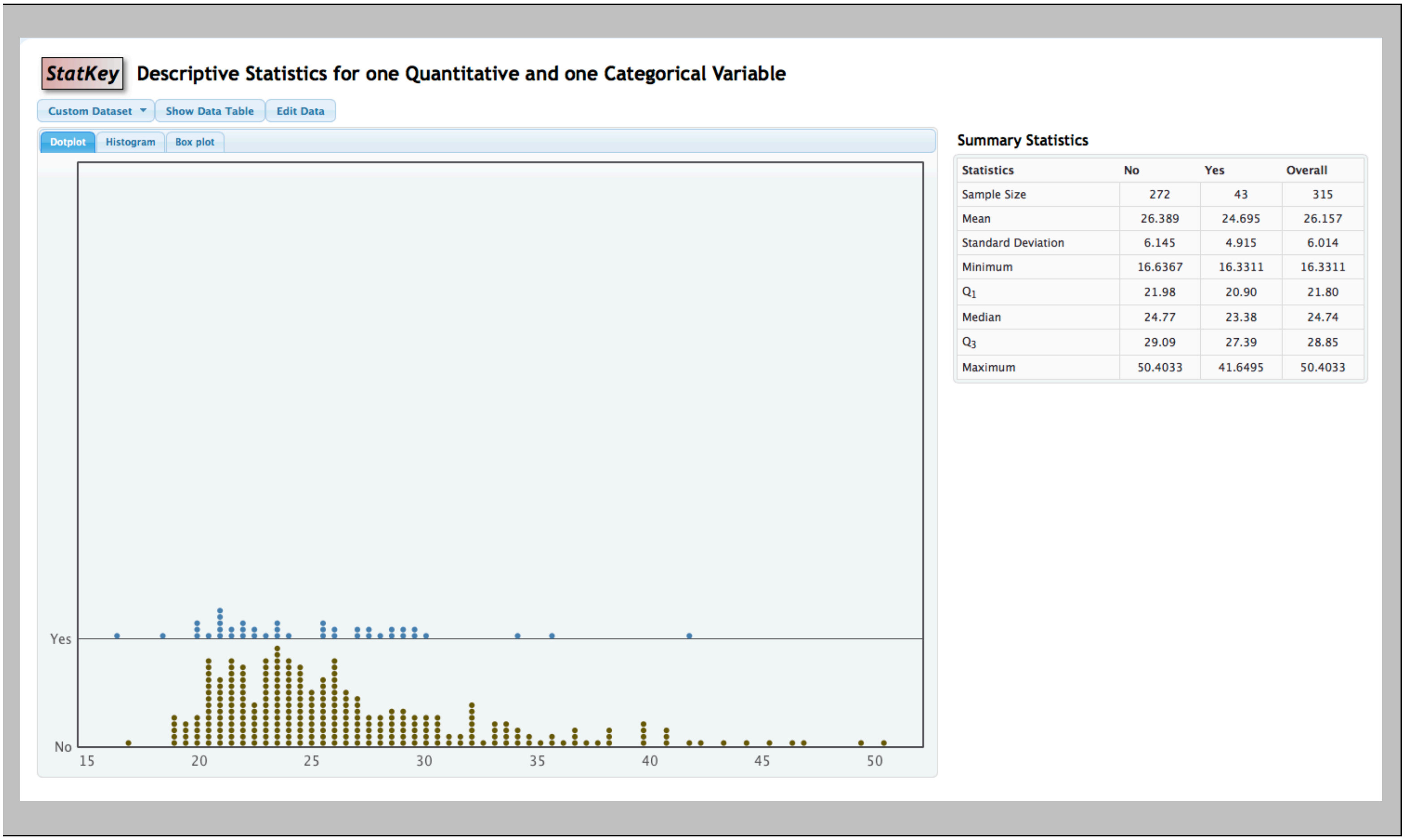


```
Smoke    Quetelet
No   21.4838
No   23.8763
No   20.0108
No   25.1406
No   20.985
No   27.5214
No   22.0115
No   28.757
No   23.0766
No   34.9699
No   20.9465
No   36.4316
No   31.7304
No   21.7885
No   27.3192
No   31.4467
No   25.9025
No   29.1526
No   38.1873
```

☑ Data has header row

Manually edit the values above or paste a tab or comma seperated file into the box and click Ok. The file must have only two columns where the first column is the categorical variable and the second is the quantitative.

"Show me": **StatKey returns an overlay dot plot summary of the distributions of the values of the <u>quantitative continuous variable</u> quetelet, separately for groups defined by the <u>categorical variable</u> smoke: smokers at the top (top), non-smokers (below). Summary statistics are provided at right: overall (n=315) and separately for non-smokers (n=272) and smokers (n=43)**



Descriptive Statistics for one Quantitative and one Categorical Variable

Summary Statistics

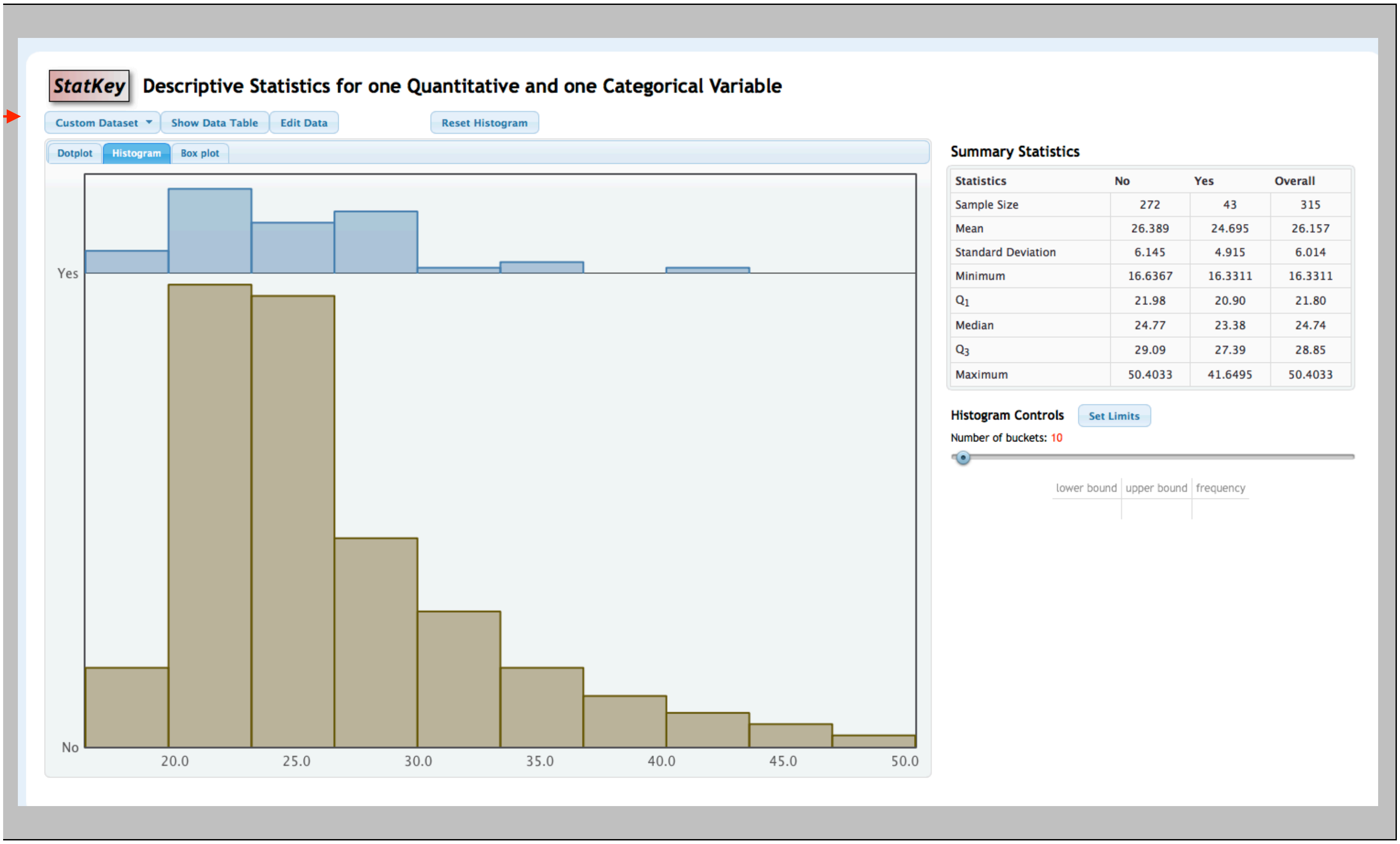
Statistics	No	Yes	Overall
Sample Size	272	43	315
Mean	26.389	24.695	26.157
Standard Deviation	6.145	4.915	6.014
Minimum	16.6367	16.3311	16.3311
Q_1	21.98	20.90	21.80
Median	24.77	23.38	24.74
Q_3	29.09	27.39	28.85
Maximum	50.4033	41.6495	50.4033

"Show me":

- **<u>Dot plots</u>** are appropriate for **<u>quantitative continuous variables</u>** (eg – quetelet).
- **<u>Overlay plots</u>** are nice in that a common axis is used, making comparison of groups defined by a categorical variable (eg – smoke) easier to interpret.
- Overall, quetelet values range from a low of 16.3 to a high of 50.4
- We see that there are very few smokers (n=43) compared to non-smokers (n=272).
- Visual comparison of the two dot plots suggests that the two distributions are similar.
- StatKey provides **<u>5 point summaries</u>**. Recall its definition: minimum, Q_1, median, Q_3, and maximum.

Click next on HISTOGRAM

StatKey now returns an overlay histogram summary of the distributions of the values of the quantitative continuous variable quetelet, separately for groups defined by the categorical variable smoke: smokers at the top (top), non-smokers (below). Again, summary statistics are provided at right: overall (n=315) and separately for non-smokers (n=272) and smokers (n=43)



Descriptive Statistics for one Quantitative and one Categorical Variable

Summary Statistics

Statistics	No	Yes	Overall
Sample Size	272	43	315
Mean	26.389	24.695	26.157
Standard Deviation	6.145	4.915	6.014
Minimum	16.6367	16.3311	16.3311
Q_1	21.98	20.90	21.80
Median	24.77	23.38	24.74
Q_3	29.09	27.39	28.85
Maximum	50.4033	41.6495	50.4033

Histogram Controls

Number of buckets: 10

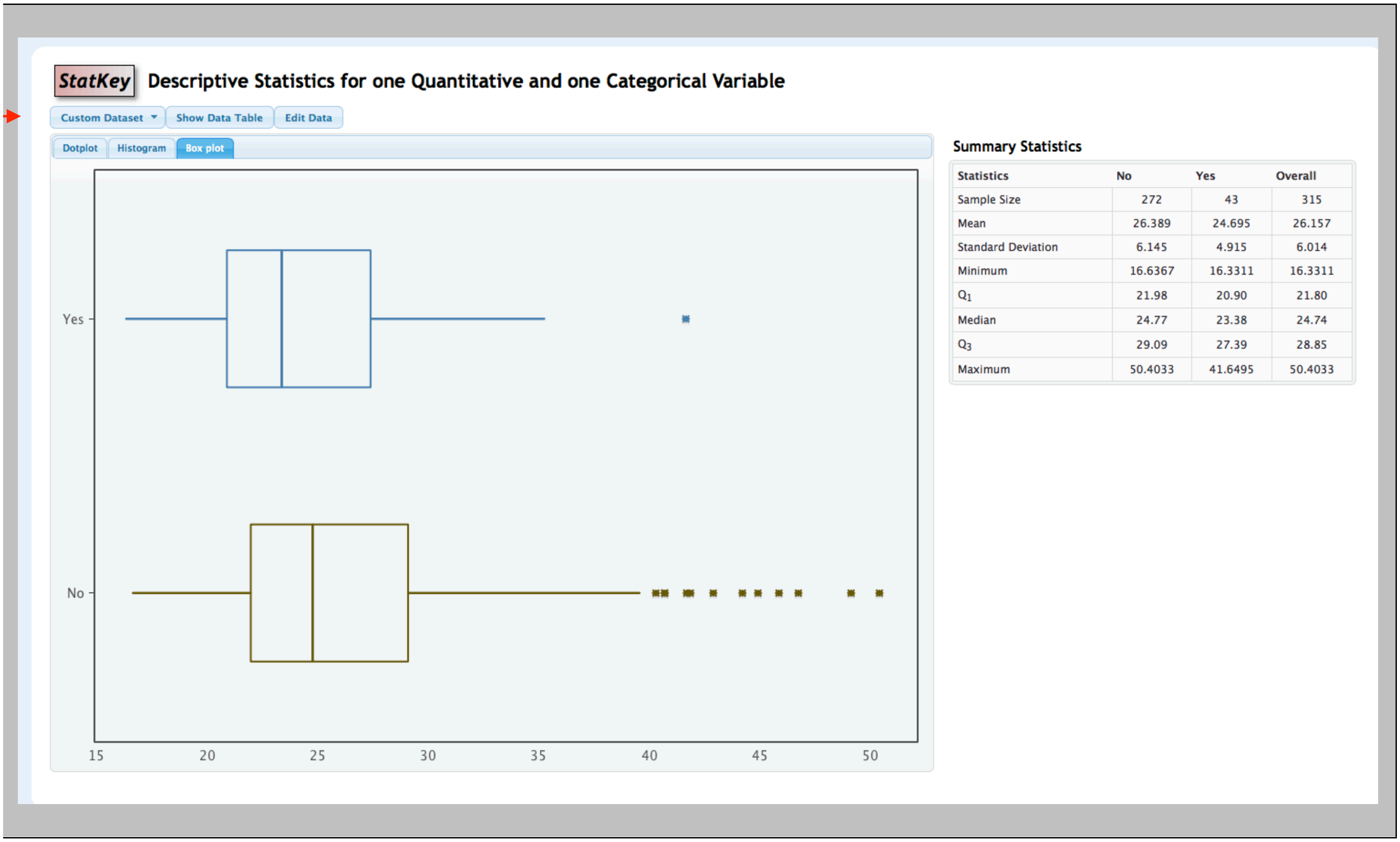
lower bound	upper bound	frequency

"Show me":

- This visual comparison makes it a little easier to appreciate that the two distributions are similar.
- ***Feeling brave?*** Click on SET LIMITS at bottom right to play with the "Histogram Controls"

Last but not least, click next on BOX PLOT

StatKey now returns an side-by-side box plot summary of the distribution of the values of the <u>quantitative continuous variable</u> quetelet, separately for groups defined by the <u>categorical variable</u> smoke: smokers at the top (top), non-smokers (below). At right summary statistics are provided: overall (n=315) separately for non-smokers (n=272) and smokers (n=43)



Descriptive Statistics for one Quantitative and one Categorical Variable

Summary Statistics

Statistics	No	Yes	Overall
Sample Size	272	43	315
Mean	26.389	24.695	26.157
Standard Deviation	6.145	4.915	6.014
Minimum	16.6367	16.3311	16.3311
Q_1	21.98	20.90	21.80
Median	24.77	23.38	24.74
Q_3	29.09	27.39	28.85
Maximum	50.4033	41.6495	50.4033

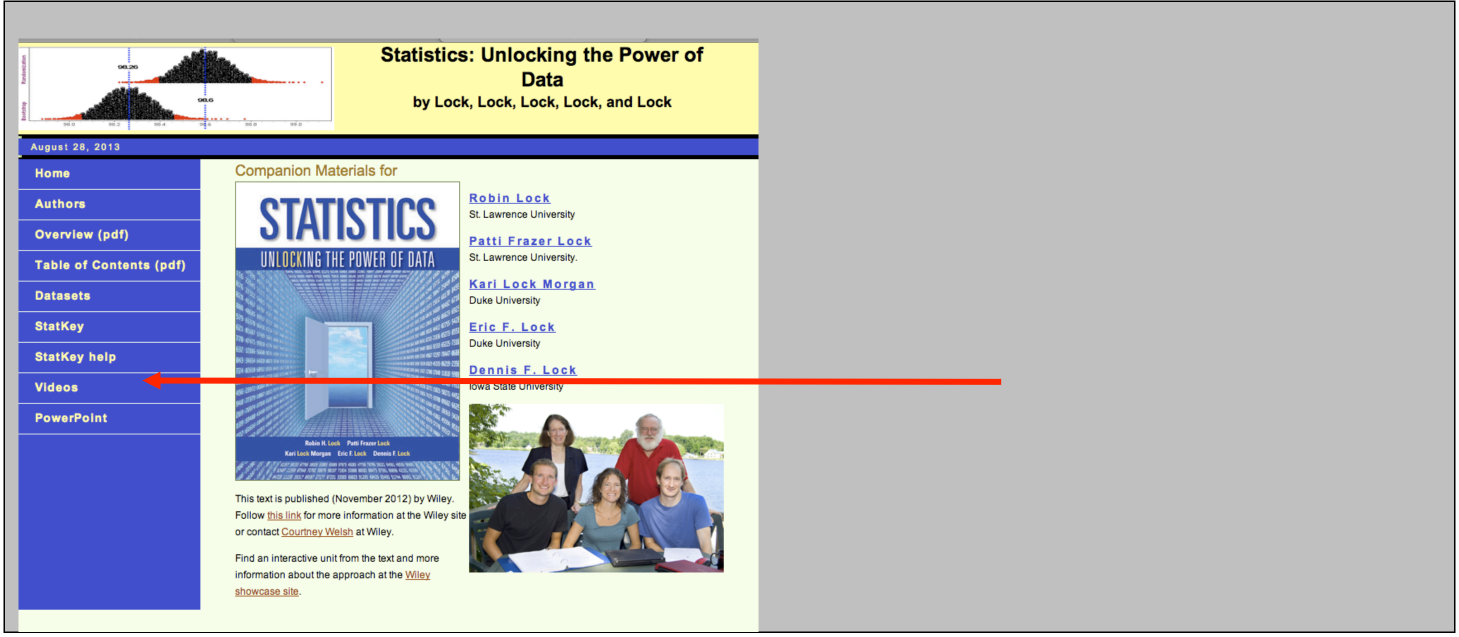
"Show me":

- My personal favorite! – cb.

Summarizing a Relationship Between Two Continuous Variables

Activity #4. Launch the StatKey tool. Then, under *Descriptive Statistics and Graphs*, click *Two Quantitative Variablese*

__1. Re-launch http://www.lock5stat.com. Then, from the menu at left, click on **StatKey:**



Statistics: Unlocking the Power of Data
by Lock, Lock, Lock, Lock, and Lock

August 28, 2013

- Home
- Authors
- Overview (pdf)
- Table of Contents (pdf)
- Datasets
- StatKey
- StatKey help
- Videos
- PowerPoint

Companion Materials for

STATISTICS
UNLOCKING THE POWER OF DATA

Robin Lock
St. Lawrence University

Patti Frazer Lock
St. Lawrence University.

Kari Lock Morgan
Duke University

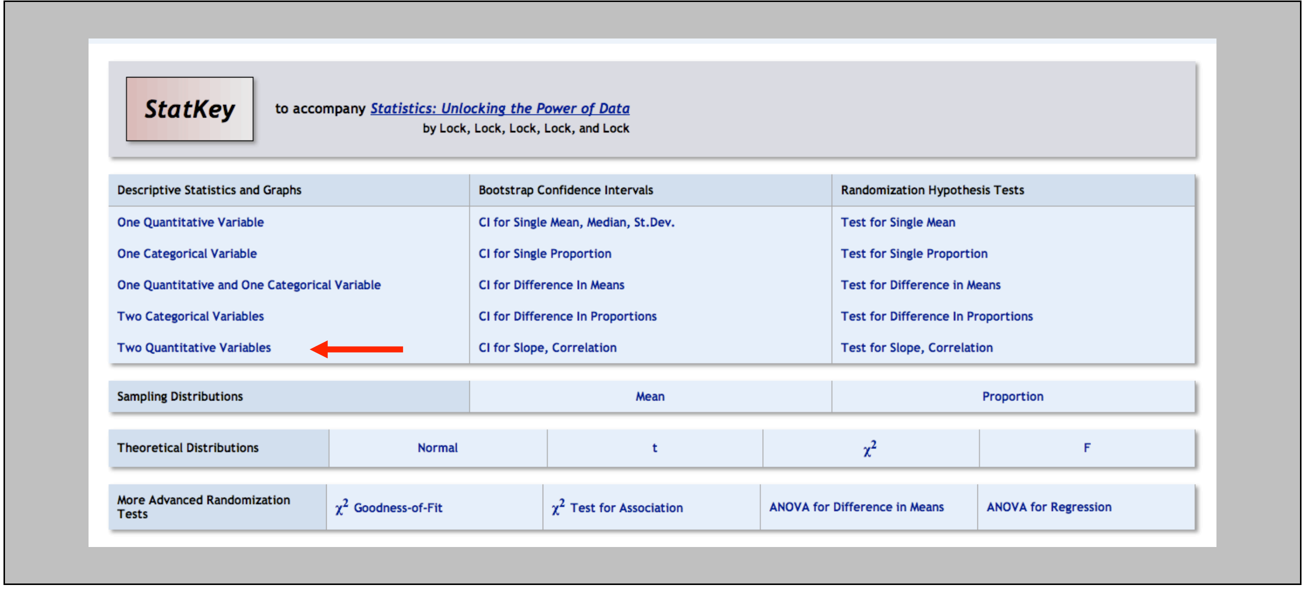
Eric F. Lock
Duke University

Dennis F. Lock
Iowa State University

This text is published (November 2012) by Wiley. Follow this link for more information at the Wiley site or contact Courtney Welsh at Wiley.

Find an interactive unit from the text and more information about the approach at the Wiley showcase site.

___2. From the selection of analysis options offered, under *Descriptive Statistics and Graphs*,
__ click ***Two Quantitative Variables***



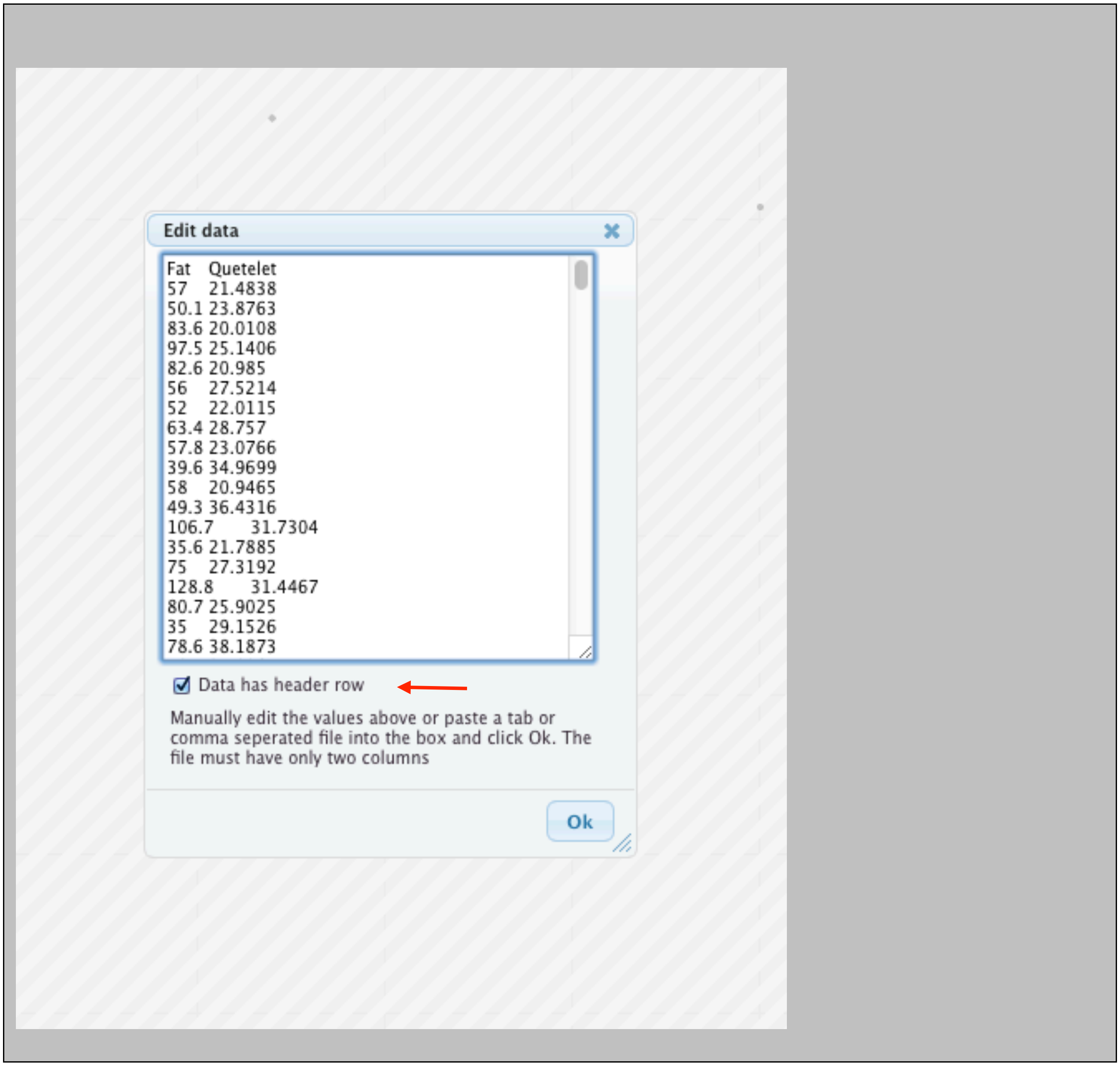
StatKey to accompany *Statistics: Unlocking the Power of Data*
by Lock, Lock, Lock, Lock, and Lock

Descriptive Statistics and Graphs	Bootstrap Confidence Intervals	Randomization Hypothesis Tests
One Quantitative Variable	CI for Single Mean, Median, St.Dev.	Test for Single Mean
One Categorical Variable	CI for Single Proportion	Test for Single Proportion
One Quantitative and One Categorical Variable	CI for Difference In Means	Test for Difference in Means
Two Categorical Variables	CI for Difference In Proportions	Test for Difference In Proportions
Two Quantitative Variables	CI for Slope, Correlation	Test for Slope, Correlation

Sampling Distributions	Mean	Proportion

Theoretical Distributions	Normal	t	χ^2	F
More Advanced Randomization Tests	χ^2 Goodness-of-Fit	χ^2 Test for Association	ANOVA for Difference in Means	ANOVA for Regression

___3. As in activity #3, replace the default data with the data on fat and quetelet.
BOX CHECK: Check to see that the box next to "Data has a header row" is checked.
__ Click on **OK** at bottom right.



Edit data

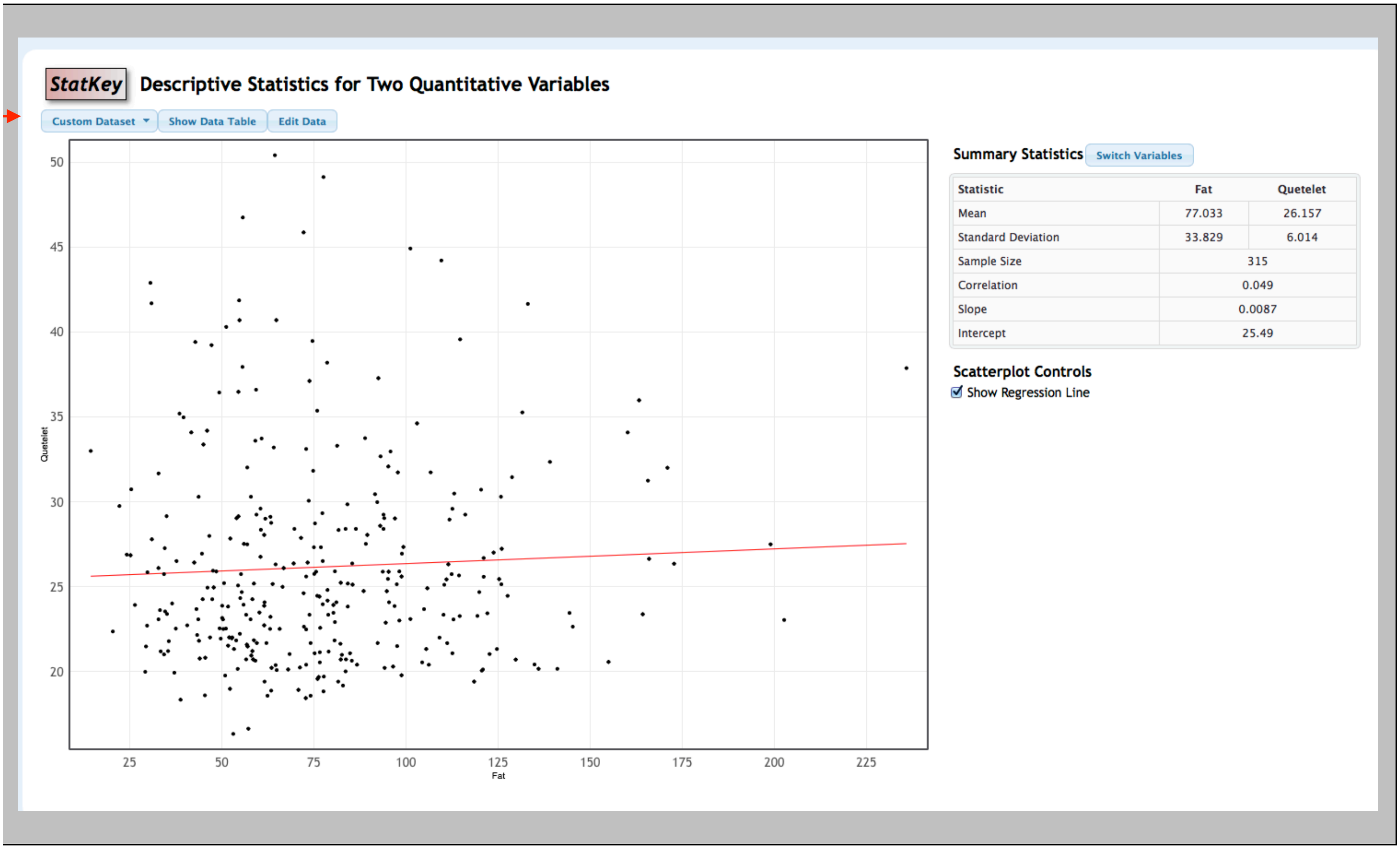
```
Fat	Quetelet
57	21.4838
50.1	23.8763
83.6	20.0108
97.5	25.1406
82.6	20.985
56	27.5214
52	22.0115
63.4	28.757
57.8	23.0766
39.6	34.9699
58	20.9465
49.3	36.4316
106.7	31.7304
35.6	21.7885
75	27.3192
128.8	31.4467
80.7	25.9025
35	29.1526
78.6	38.1873
```

☑ Data has header row

Manually edit the values above or paste a tab or comma seperated file into the box and click Ok. The file must have only two columns

StatKey returns a scatterplot summary of the co-variation in the values of fat (plotted on the “horizontal x-axis”). Selected summary statistics are shown at right.

Click on the box “Show Regression Line” at right, under “Scatterplot Controls”



Summary Statistics

Statistic	Fat	Quetelet
Mean	77.033	26.157
Standard Deviation	33.829	6.014
Sample Size	315	
Correlation	0.049	
Slope	0.0087	
Intercept	25.49	

Scatterplot Controls

☑ Show Regression Line

“Show me”:

- In this sample, most of the data are in the region where FAT < 150 and QUETELET < 35.
- Data values outside this region occur but are rare.
- The regression line is an example of **simple linear regression**, in which Y=quetelet is modeled linearly in X=fat.
- Slope = 0.0087 says that associated with a one gram increase in fat, is an estimated 0.0087 increase in quetelet.
- Intercept = 25.49 says that when fat=0, it is estimated that quetelet=25.49. This illustrates the general finding that the validity of a linear relationship is limited and does not extend to being valid at X=0.
- The fitted line looks rather flat, suggesting that the relationship between fat and quetelet is weak. A weak relationship means that the predictor (in this case fat) is not a very good predictor of outcome (in this case quetelet). Another way of saying this is to say that fat is not useful in explaining the variability in quetelet.