


# SPPH567 TUTORIAL II

# TODAY'S TASKS

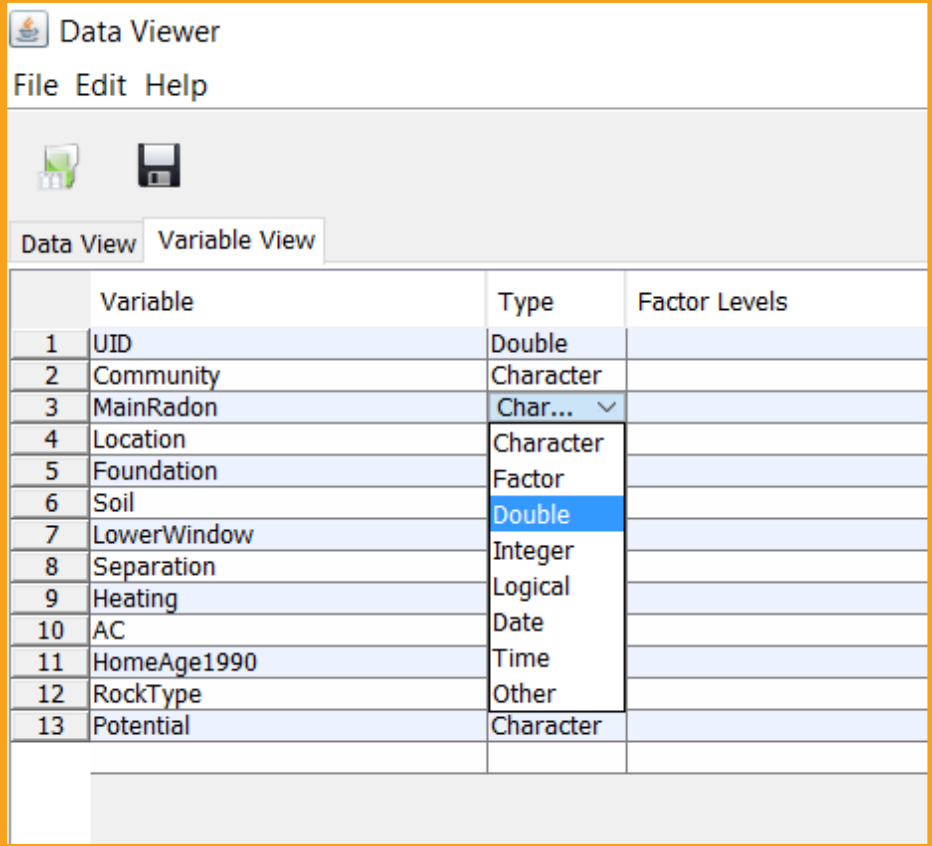
- Make changes to the data
- Calculate summary statistics
- Make frequency distribution plot (histogram)
- Make Q-Q plot
- Do goodness of fit test

# CHANGE DATA TYPE AND VARIABLE NAME

- Go to Variable View in the Data Viewer
- Click on the cell in the column Type for the variable you want to change
- Select the correct type in the drop-down menu.
- You can also change the name of the variable by double clicking the variable name

Categorical,  
dichotomous,  Factor  
Binary

Continuous  Double, Integer



The screenshot shows the SPSS Data Viewer window in Variable View. The window title is "Data Viewer" and it has a menu bar with "File", "Edit", and "Help". Below the menu bar are icons for a document and a floppy disk. The "Variable View" tab is selected, showing a table with columns for "Variable", "Type", and "Factor Levels". The table lists 13 variables. The "Type" column for the variable "Soil" (row 6) is currently set to "Double" and is highlighted in blue. A dropdown menu is open for the "MainRadon" variable (row 3), showing a list of data types: "Character", "Integer", "Logical", "Date", "Time", "Other", and "Character".

	Variable	Type	Factor Levels
1	UID	Double	
2	Community	Character	
3	MainRadon	Char... ▾	
4	Location	Character	
5	Foundation	Factor	
6	Soil	Double	
7	LowerWindow	Integer	
8	Separation	Logical	
9	Heating	Date	
10	AC	Time	
11	HomeAge1990	Other	
12	RockType	Character	
13	Potential	Character	

# CHANGE THE ORDER OF LEVELS FOR CATEGORICAL VARIABLES

1. Click on the Factor levels column

2. Change the order of the levels up/down

Variable	Type	Factor Levels
1 UID	Double	
2 Community	Factor	(1) ATLIN; (2) BARRIERE; (3) BLUEBERRY; (4) BR...
3 MainRadon	Double	
4 Location	Factor	(1) Bedroom; (2) Crawl; (3) Kitchen; (4) Laundry...
5 Foundation	Factor	(1) NA; (2) Other; (3) Poured;
6 Soil	Factor	(1) Clay; (2) Loam; (3) NA;
7 LowerWindow	Factor	(1) Closed; (2) NA; (3) Open;
8 Separation	Factor	(1) Door; (2) NA; (3) No Door;
9 Heating	Factor	(1) Electric; (2) Forced Air; (3) Hot Water; (4) NA...
10 AC	Factor	(1) NA; (2) No; (3) Yes;
11 HomeAge1990	Integer	
12 RockType	Factor	(1) Metamorphic; (2) Plutonic; (3) Sedimentary; (...)
13 Potential	Factor	(1) HIGH; (2) LOW; (3) MOD;

Factor Editor: Potential

Levels

LOW

MOD

HIGH

Contrasts  Ordered

OK Cancel

# DROP/ADD A LEVEL FOR CATEGORICAL VARIABLES

The screenshot shows the SPSS Data Viewer interface with the 'Variable View' tab selected. The 'Soil' variable is highlighted in blue. The 'Factor Levels' column for 'Soil' is '(1) Clay; (2) Loam; (3) NA;'. The 'Factor Editor: Soil' dialog box is open, showing the 'Levels' list with 'Clay', 'Loam', and 'NA'. The 'NA' level is selected. The dialog box has a red minus sign button next to the 'NA' level, which is being pointed to by an orange callout bubble. Another orange callout bubble points to the 'Factor Levels' column in the main window. A third orange callout bubble points to the plus sign button in the dialog box.

Variable	Type	Factor Levels
1 UID	Double	
2 Community	Factor	(1) ATLIN; (2) BARRIERE; (3) BLUEBERRY; (4) BR...
3 MainRadon	Double	
4 Location	Factor	(1) Bedroom; (2) Crawl; (3) Kitchen; (4) Laundry...
5 Foundation	Factor	(1) Other; (2) Poured;
6 Soil	Factor	(1) Clay; (2) Loam; (3) NA;
7 LowerWindow	Factor	(1) Closed; (2) NA; (3) Open;
8 Separation	Factor	(1) Door; (2) NA; (3) No Door;
9 Heating	Factor	(1) Electric; (2) Forced Air; (3) Hot Water; (4) NA...
10 AC	Factor	(1) No; (2) Yes;
11 HomeAge1990	Integer	
12 RockType	Factor	(1) Metamorphic; (2) Plutonic; (3) Sedimentary; (...)
13 Potential	Factor	(1) LOW; (2) MEDIUM; (3) HIGH;

NA should not be a level, so drop it

1. Click on the Factor levels column

2. Drop or add a level with +/-



# LOG TRANSFORM

1. Go to Data > Transform

2. Select MainRadon here

3. Click the arrow and you should see the variable added here

4. Select log(x) from this drop-down menu

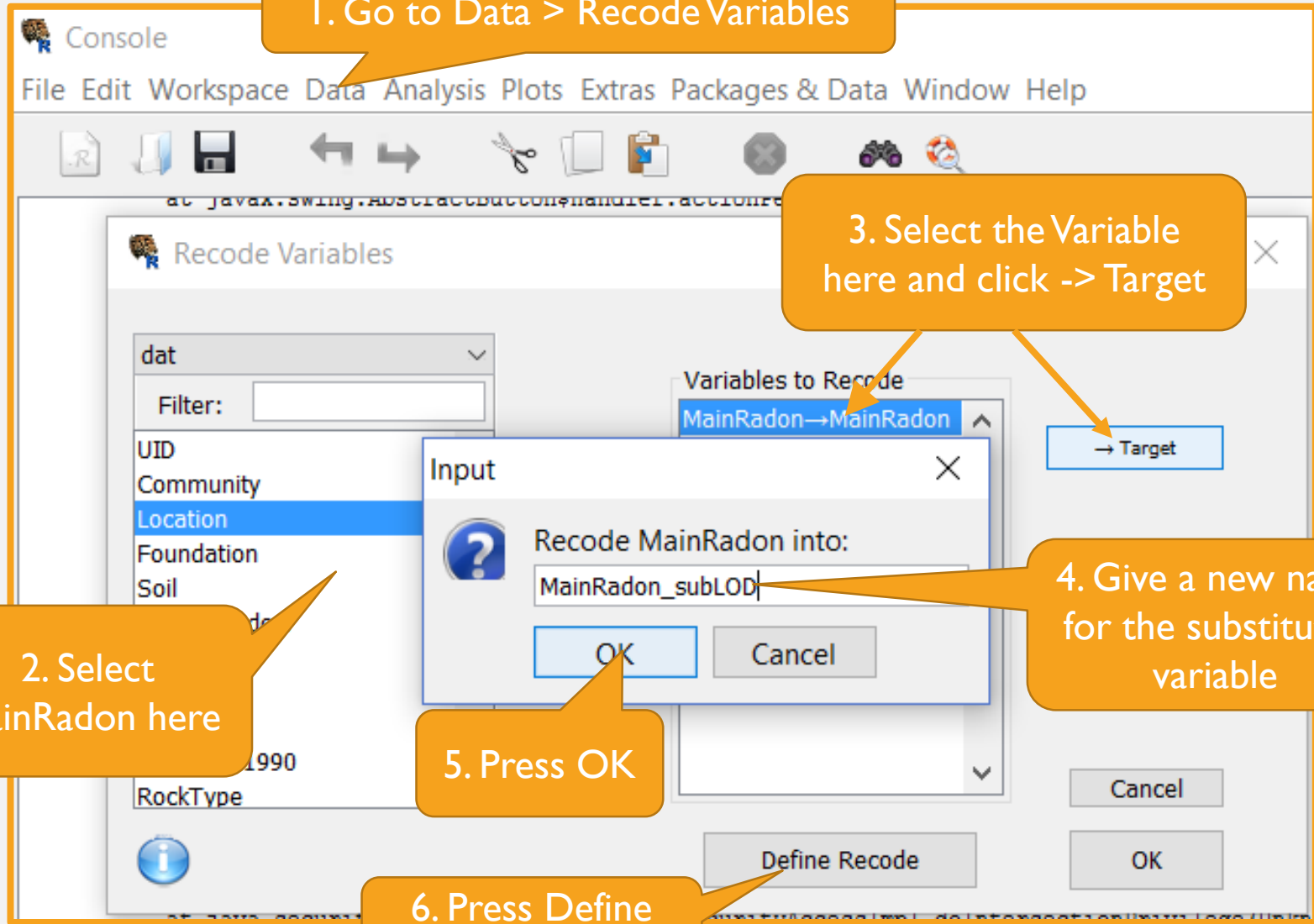
5. Press Run

The screenshot shows the RStudio interface with the 'Transform Variables' dialog box open. The 'dat' data frame is selected, and the 'Location' variable is highlighted in the list. The 'MainRadon' variable is moved to the 'Variables to Transform' list. The 'Transformation' dropdown is set to 'Natural log: log(x)'. The 'Run' button is highlighted.

```
> library(XLConnect)
Loading required packages:
X
> library(XLConnect)
> dat <- readWorksheet(loadWorkbook('G:/cou
> dat[,2]<-as.factor(dat[,2])
> dat[,3]<-as.double(dat[,3])
Warning message:
NAs introduced by coercion
> dat[,4]<-as
> dat[,5]<-as
> dat[,6]<-as
> dat[,7]<-as
> dat[,8]<-as
> dat[,9]<-as.factor(dat[,9])
> dat[,10]<-as.factor(dat[,10])
> dat[,11]<-as.integer(dat[,11])
> dat[,12]<-as.factor(dat[,12])
> dat[,13]<-as.factor(dat[,13])
> dat[,13]<-factor(dat[,13], levels=c("LOW",
> dat[,10]<-factor(dat[,10], levels=c("No",
> dat[,5]<-factor(dat[,5], levels=c("Other",
> dat[,6]<-factor(dat[,6], levels=c("Clay", "Loam", NA ), order
```

# SUBSTITUTE <LOD VALUES – I

1. Go to Data > Recode Variables



2. Select MainRadon here

3. Select the Variable here and click -> Target

4. Give a new name for the substituted variable

5. Press OK

6. Press Define Recode



# SUBSTITUTE <LOD VALUES – 2

1. Click on the variable

2. You should see a summary of distribution here

percentiles	
0%	15.10
10%	19.99
20%	24.90
30%	32.07
40%	39.36
50%	49.85
60%	63.76
70%	83.76
80%	112.16
90%	179.36
100%	1707.20

3. Setup here to define values of NA to the value for substitution (you need to put in the exact number for substitution)

4. Press Add and you should see a line of code under Recodings

5. Press OK

# SUMMARY STATISTICS -- FREQUENCIES

1. Go to Analysis > Frequencies

2. Select the variables you want to run frequencies on

3. Click the arrow and you should see the selected variables here

4. Press OK

The screenshot shows the SPSS 'Run Frequencies' dialog box. The 'dat' dataset is selected. The 'Filter' field is empty. The list of variables includes UID, Community, MainRadon, Location, Soil, Separation, AC, HomeAge1990, RockType, and Potential. The 'Run Frequencies On:' list contains Foundation, LowerWindow, and Heating. The 'OK' button is highlighted.

Value	# of Cases
1 Other	159
2 Poured	964

# SUMMARY STATISTICS – FREQUENCIES (CON'T)

```
> frequencies(dat[c("Potential")] , r.digits = 1)
```

```
$Potential
```

```
-----  
--                               Frequencies                               --  
--                               -----                               --  
Value # of Cases % Cumulative %  
1  LOW      422   37.2      37.2  
2  MOD      290   25.6      62.8  
3  HIGH      422   37.2     100.0  
--                               -----                               --  
--                               Case Summary                               --  
--                               -----                               --  
Valid Missing Total  
# of cases 1134    0 1134  
--                               -----                               --  
--                               -----                               --  
-----
```

In the Console window, you should see a table like this

Frequencies – number of cases

Proportion of the total number of valid cases

Cumulative proportion

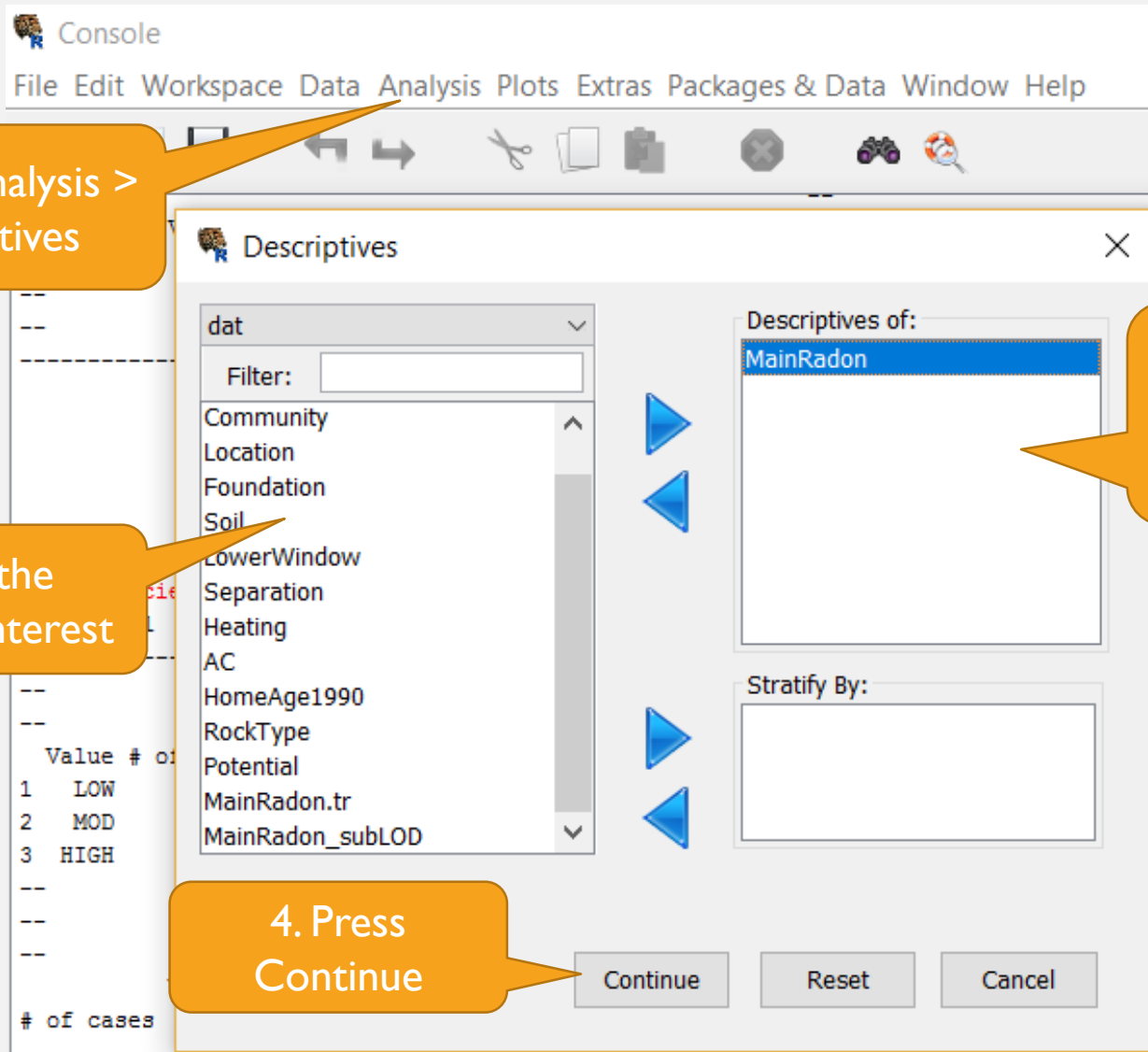
# SUMMARY STATISTICS – DESCRIPTIVES

1. Go to Analysis > Descriptives

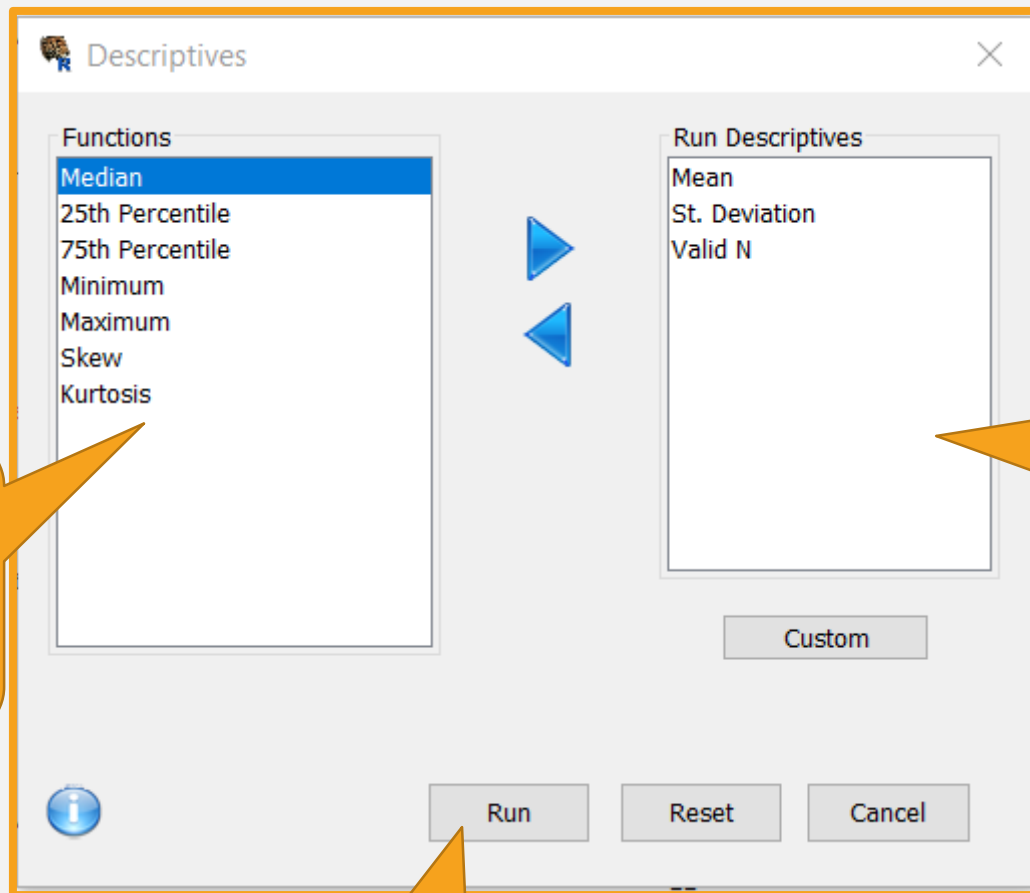
2. Select the variable of interest

3. Click the arrow to put the selected variables here

4. Press Continue



# SUMMARY STATISTICS – DESCRIPTIVES (CON'T)



1. Select the descriptive statistics you want to calculate

2. Click the arrow to put the selected variables here

3. Press Run

# MAKE HISTOGRAM - I

1. Go to Plots > Plot Builder

2. Click the tab Geometric Elements

3. Double click histogram

The screenshot shows the R Studio Plot Builder window. The 'Geometric Elements' tab is active, and the 'histogram' option is highlighted in blue. The console on the left shows the following R code and output:

```
> frequencies(dat[c("Poten", "Strata"), "Potential"])
# A tibble: 3 x 4
  Value # of Cases %
1 LOW 422 37.2
2 MOD 290 25.6
3 HIGH 422 37.2

Valid Missing T
# of cases 1134 0

> descriptive.table(vars = c("Poten", "Strata"),
+ func.names = c("Mean", "St", "sd", "var", "min", "max"),
+ strata: all cases `
  Mean.MainRadon St
  89.83533
```

# MAKE HISTOGRAM - 2

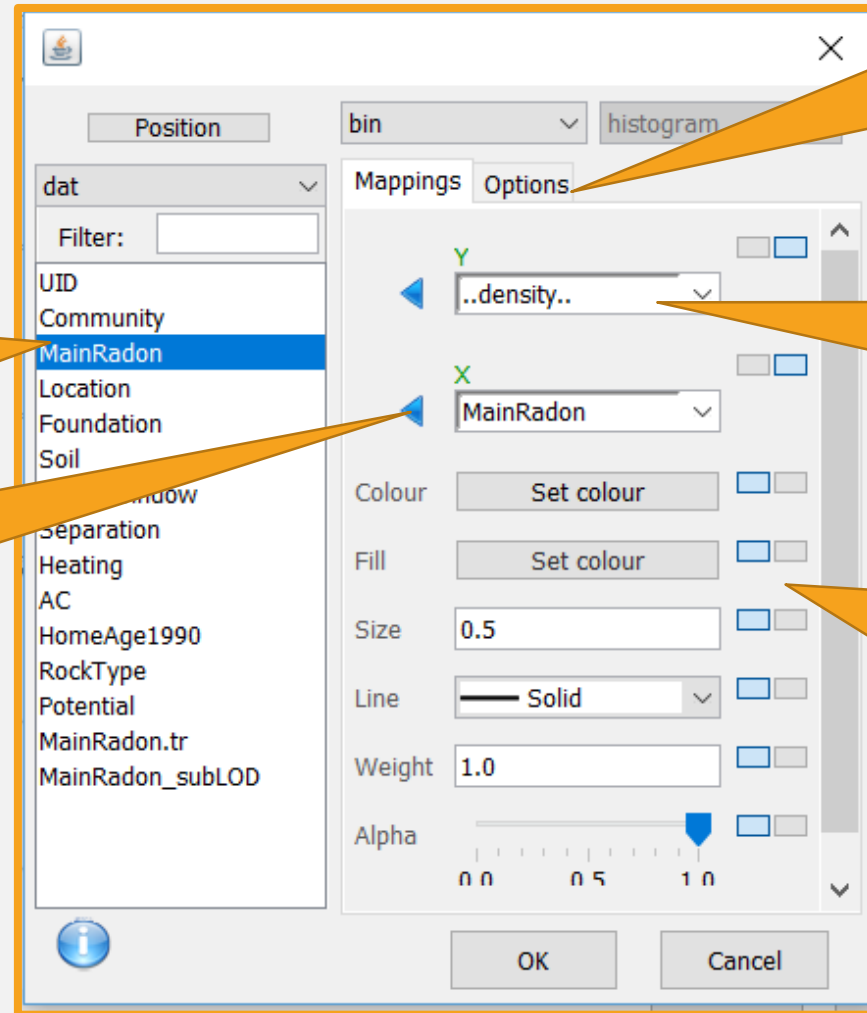
1. Select the variable to put at X axis

2. Click this arrow to add the variable to X

5. You can change the bin width in the Options

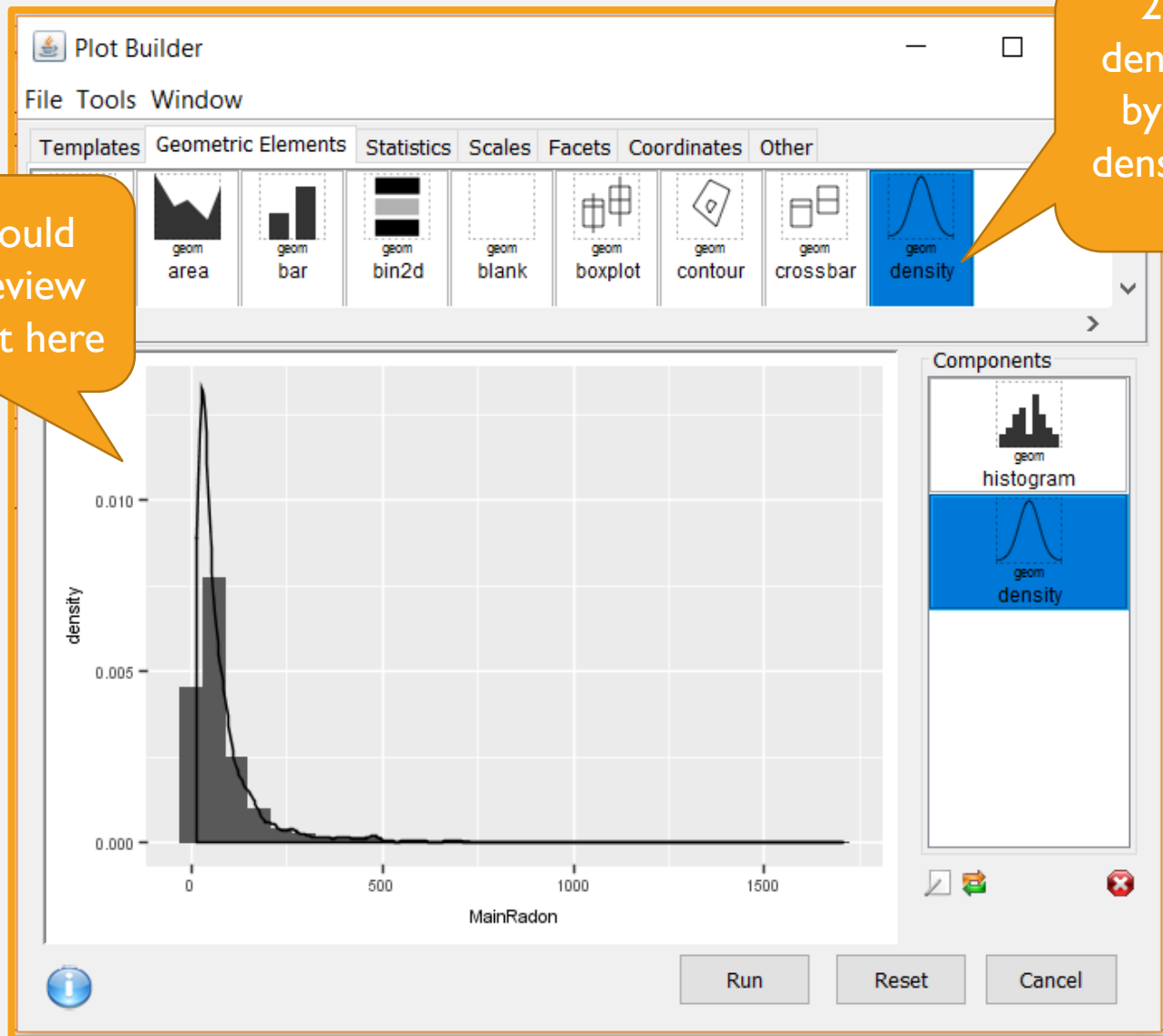
3. Select density or count for Y

4. You can change the color/fill or other graphical setting



# MAKE HISTOGRAM - 3

1. You would see a preview of the plot here



2. You can add a density function line by double clicking density in Geometric Elements



# CHANGE X OR Y LABELS

The screenshot shows the Plot Builder interface with a histogram plot. The x-axis is labeled 'MainRadon' and the y-axis is labeled 'count'. The plot shows a distribution of MainRadon values with a peak count of approximately 280. The 'Other' tab is selected in the top menu, and the 'y lab' theme is highlighted in the theme selection area. The 'Run' button is highlighted in the bottom right corner.

1. Click tab Other

2. Double click xlab or ylab to change the label

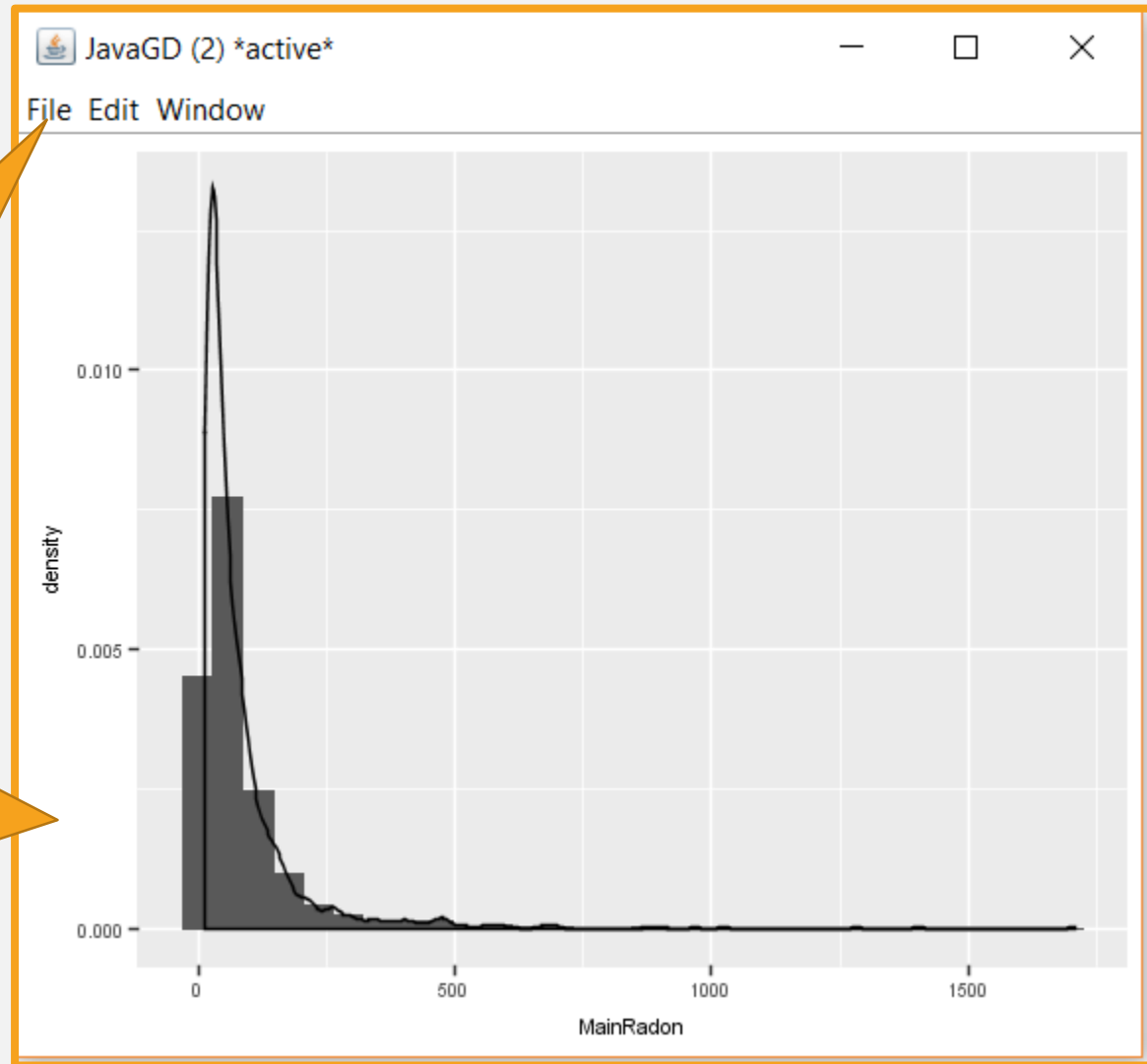
3. You can also add a title to the plot here

4. Press Run when the plot looks good in preview

# MAKE HISTOGRAM - 4

2. To save the plot, go to File > save as any format in the options

1. Now the plot is made and you can change the size by dragging the window

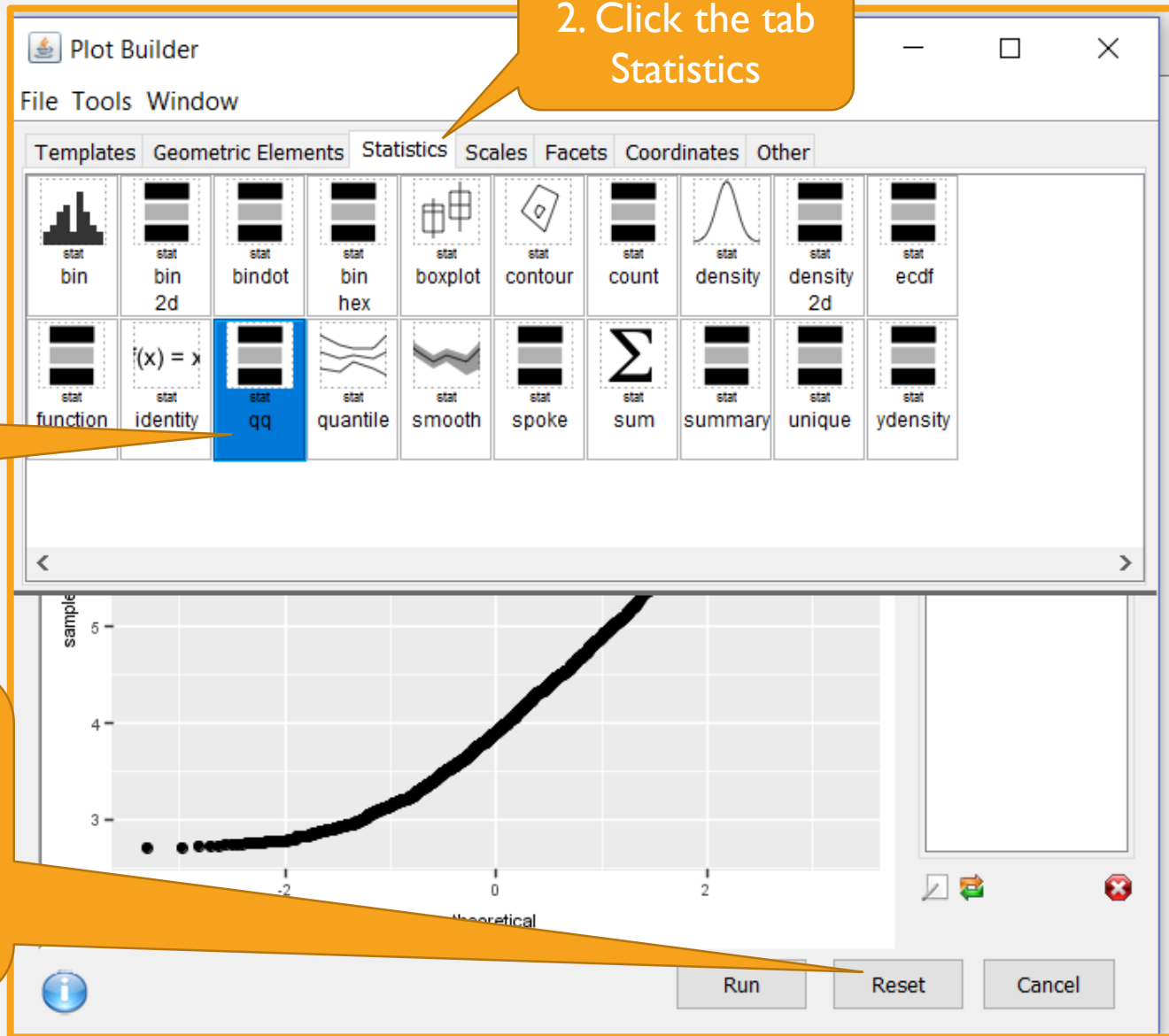


# MAKE Q-Q PLOT - I

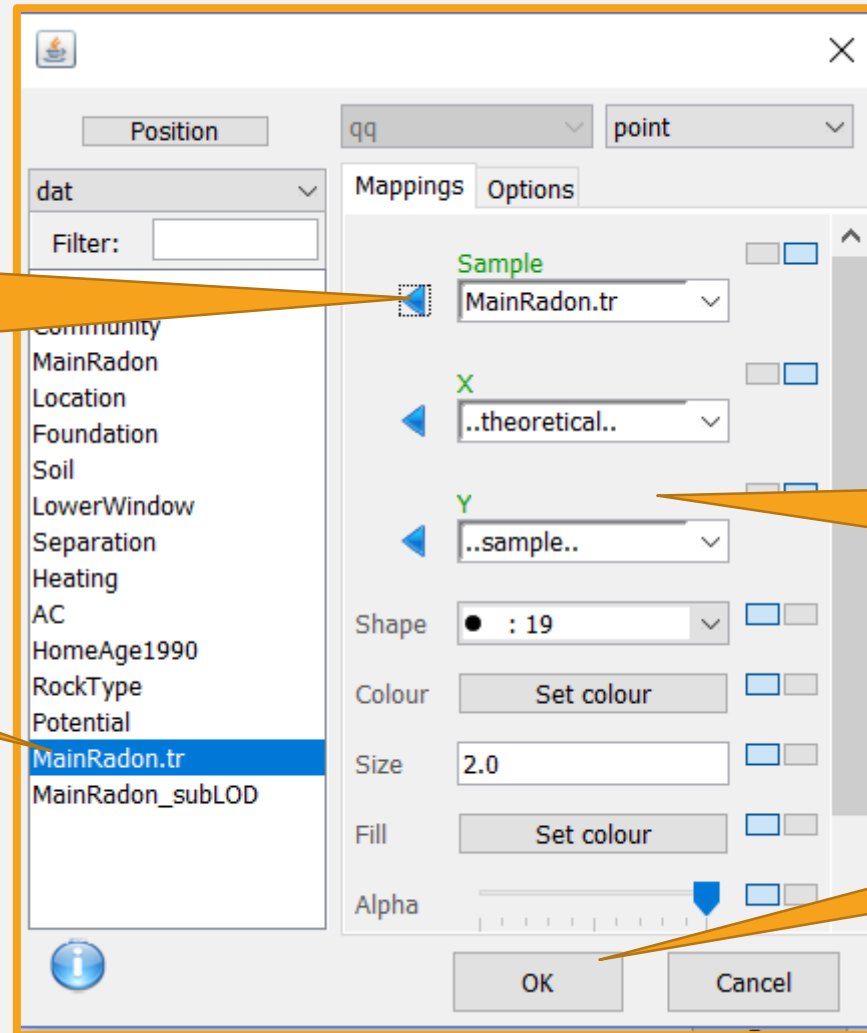
2. Click the tab Statistics

3. Double click qq

1. Go to Plot Builder and press Reset to remove any components from the last plot



# MAKE Q-Q PLOT - 2



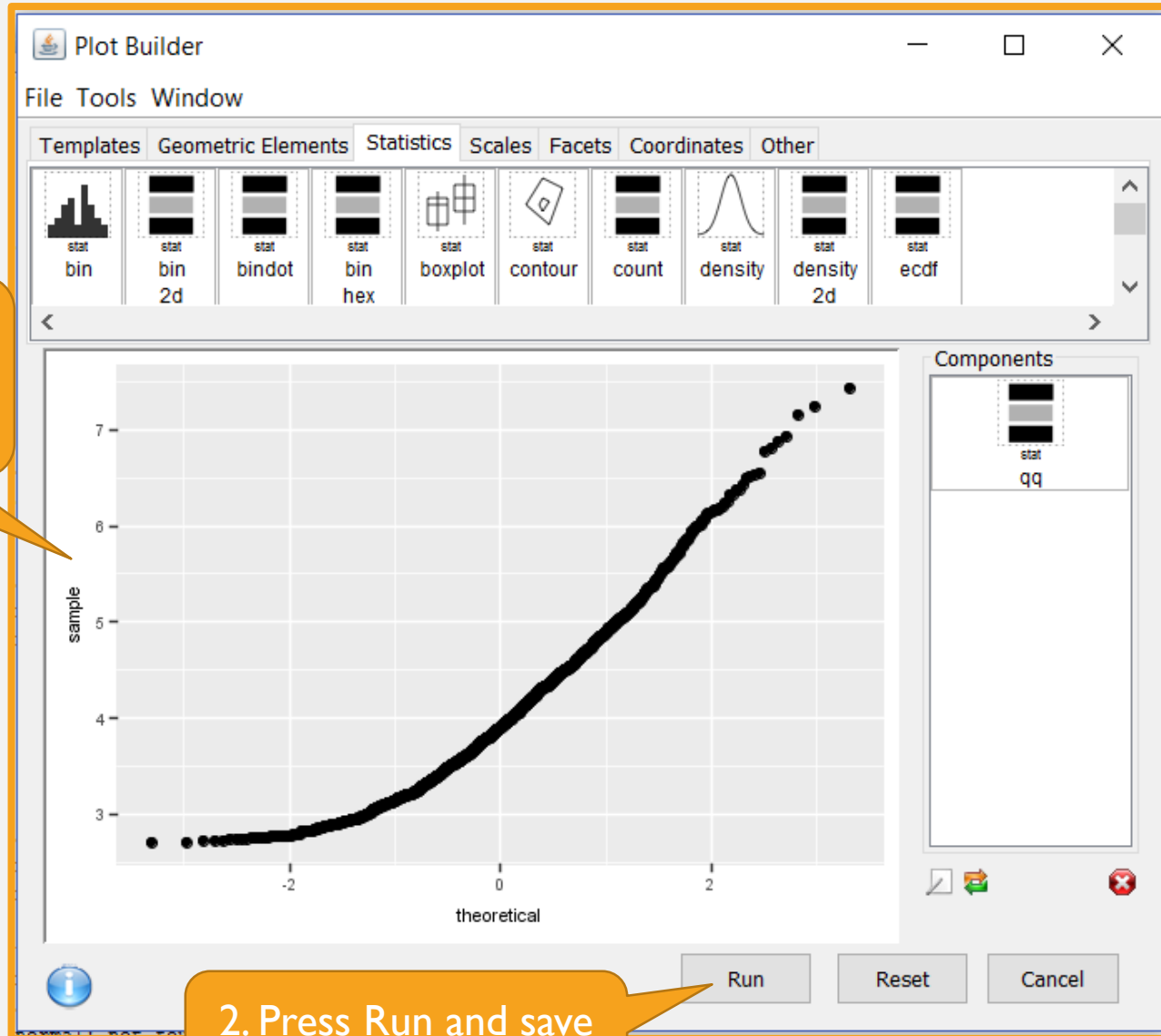
2. Click the arrow here to add the variable to Sample

1. Select the variable to plot

3. Make sure X is theoretical, and Y is sample

4. Press OK

# MAKE Q-Q PLOT - 3



1. Now we have the preview of the qq plot

2. Press Run and save the plot as in page 16

# GOODNESS OF FIT TEST

1. Go to Analysis > One Sample Test

2. Select the variable(s) for the normality test

4. Check Shapiro-Wilk test against normality

3. Add them here by clicking the arrow

5. Press Run and the results should show up in the Console

The screenshot shows the R Studio interface with the 'One Sample Test' dialog box open. The 'Data' field is set to 'dat'. The 'Variables' list on the right contains 'MainRadon' and 'MainRadon.tr'. The 'Tests' section has the 'Shapiro-Wilk test against normality' checked. The 'Run' button is highlighted. The console window in the background shows the following code and output:

```
> one.sa
+ data=c
+ test=shapiro.test
```

The console also displays a warning message: 'Warning: 1: Removal of some variables caused by missing data: MainRadon, MainRadon.tr'.