

SPPH567 TUTORIAL VI

TODAY'S TASKS

- Subset the data
- Recode continuous variable
- Make grouped bar plots
- Cross-tabulation and Chi-squared tests
- Do logistic regression

SUBSET DATA -- I

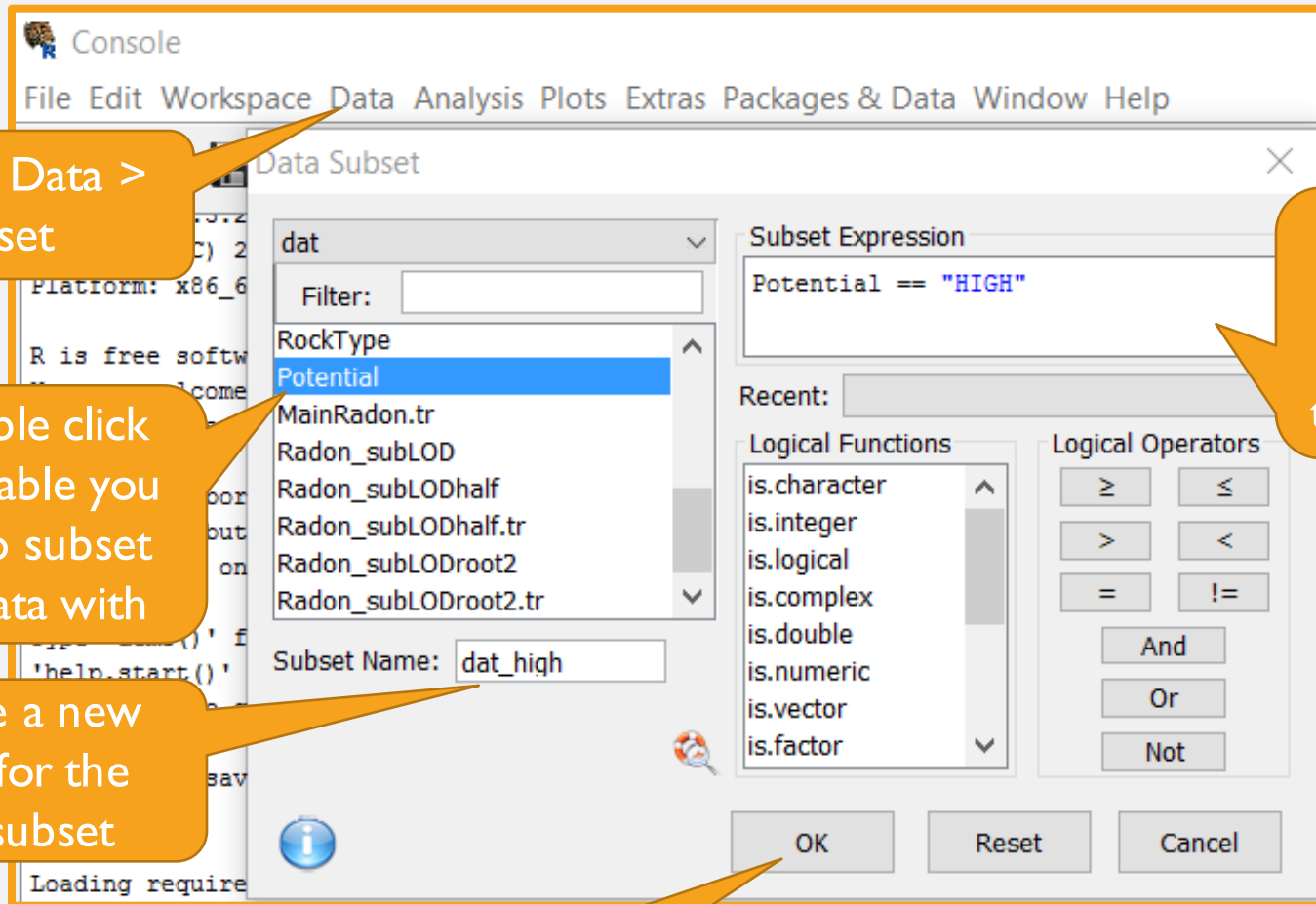
1. Go to Data > Subset

2. Double click the variable you want to subset your data with

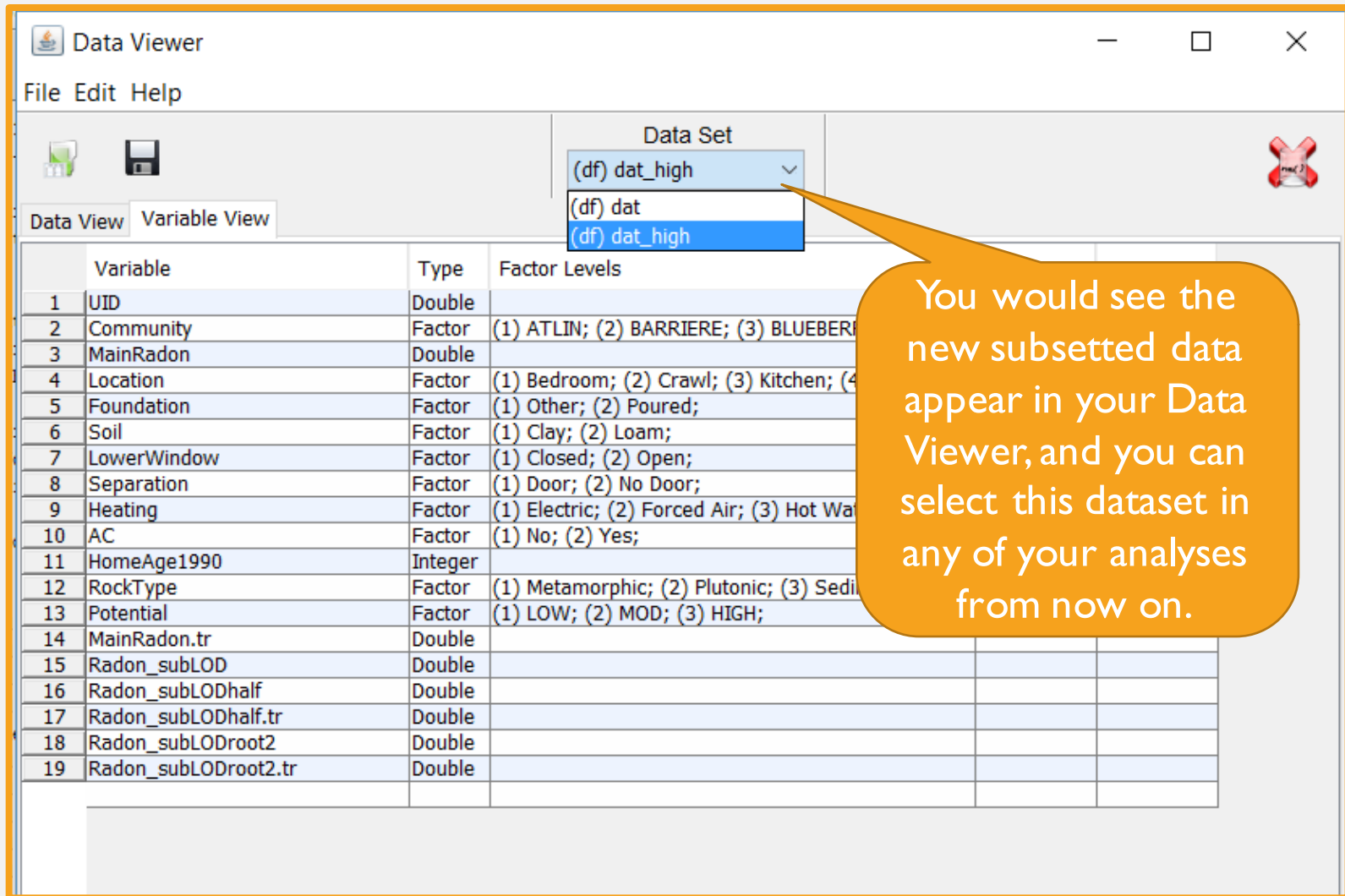
3. Make a new name for the data subset

4. Type in or use the operators below to define the subset criteria

5. Click OK



SUBSET DATA -- 2



The screenshot shows the SPSS Data Viewer window. The 'Data Set' dropdown menu is open, showing three options: '(df) dat_high' (selected), '(df) dat', and '(df) dat_high'. An orange callout bubble points to the selected option with the text: "You would see the new subsetted data appear in your Data Viewer, and you can select this dataset in any of your analyses from now on."

	Variable	Type	Factor Levels
1	UID	Double	
2	Community	Factor	(1) ATLIN; (2) BARRIERE; (3) BLUEBERG
3	MainRadon	Double	
4	Location	Factor	(1) Bedroom; (2) Crawl; (3) Kitchen; (4)
5	Foundation	Factor	(1) Other; (2) Poured;
6	Soil	Factor	(1) Clay; (2) Loam;
7	LowerWindow	Factor	(1) Closed; (2) Open;
8	Separation	Factor	(1) Door; (2) No Door;
9	Heating	Factor	(1) Electric; (2) Forced Air; (3) Hot Wa
10	AC	Factor	(1) No; (2) Yes;
11	HomeAge1990	Integer	
12	RockType	Factor	(1) Metamorphic; (2) Plutonic; (3) Sedi
13	Potential	Factor	(1) LOW; (2) MOD; (3) HIGH;
14	MainRadon.tr	Double	
15	Radon_subLOD	Double	
16	Radon_subLODhalf	Double	
17	Radon_subLODhalf.tr	Double	
18	Radon_subLODroot2	Double	
19	Radon_subLODroot2.tr	Double	

SUBSET DATA -- 3

Useful expressions for subsetting data

Potential == "HIGH" Equals

Potential != "HIGH" Not equal to

HomeAge < 20 & Potential == "HIGH" And

HomeAge < 20 | Potential == "HIGH" Or

Potential %in% c("LOW", "MOD") Is an element of

Note: values in factor or character variables need to be quoted

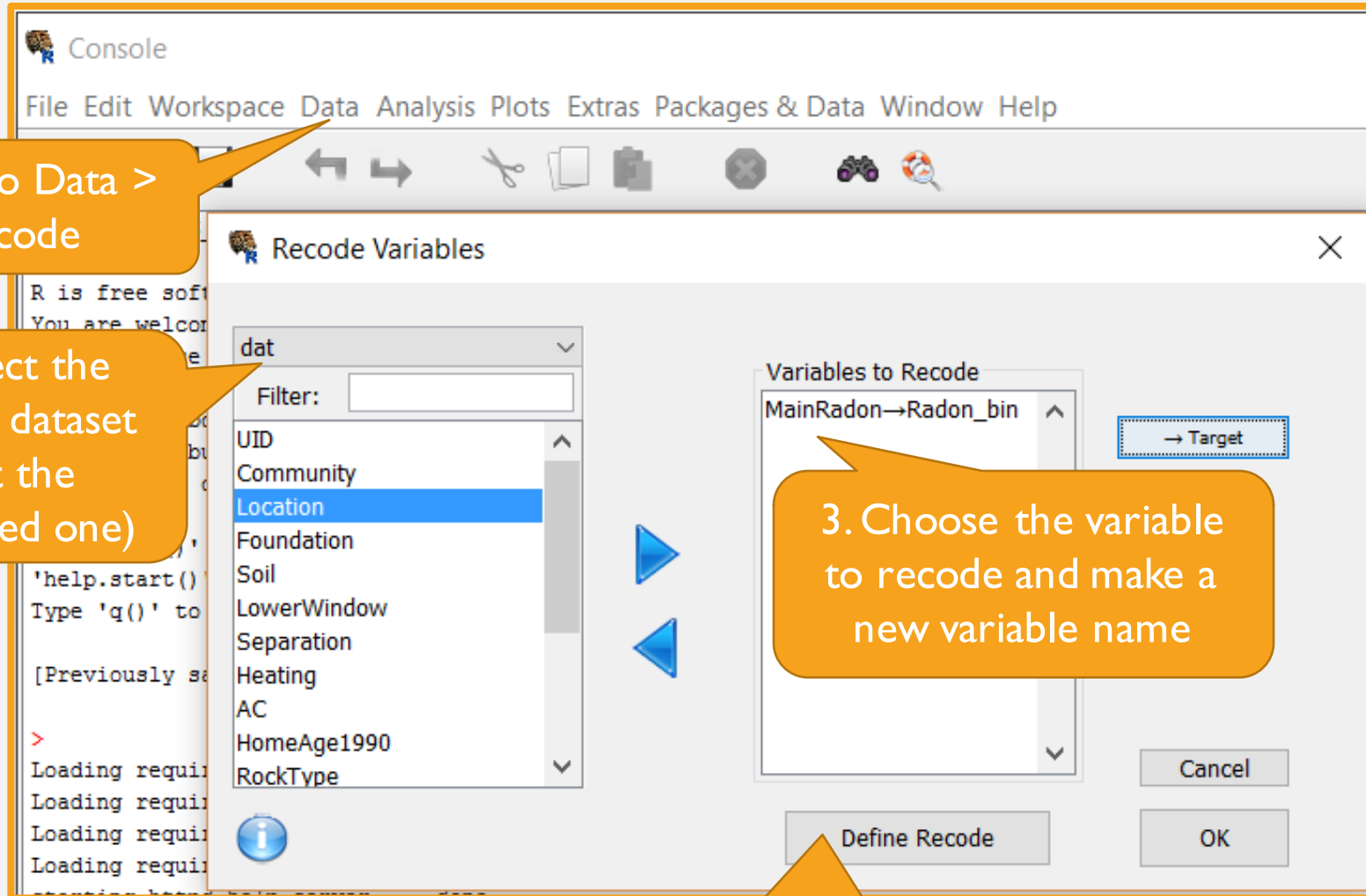
RECODE CONTINUOUS VARIABLE -- I

1. Go to Data > Recode

2. Select the original dataset (not the subsetted one)

3. Choose the variable to recode and make a new variable name

4. Click Define Recode



RECODE CONTINUOUS VARIABLE -- 2

The screenshot shows the 'Set Variable Codings' dialog box in SPSS. The 'Variable Information' section shows 'MainRadon' being recoded into 'Radon_bin'. The 'Code' section is configured with the following settings:

- Value \geq 100 into 1 (Add)
- Value \leq [] into [] (Add)
- Else: 0 (Add)

The 'Recodings' section displays the resulting mapping: '100:Hi -> 1' and 'else -> 0'. The 'percentiles' table is also visible.

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

Four instructional callouts are present:

1. Set values ≥ 100 as 1, click Add
2. Set all other values as 0, click Add
3. You should see the recoding criteria here
4. Click OK

RECODE CONTINUOUS VARIABLE -- 2

The screenshot shows the 'Set Variable Codings' dialog box in SPSS. The 'Variable Information' section shows 'MainRadon' being recoded into 'Radon_bin'. The 'Code' section has three rows: the first row is set to 'Value ≥ 100 into 1', the second row is empty, and the third row is 'Else: 0'. The 'Recodings' section shows '100:Hi -> 1' and 'else -> 0'. Four orange callout boxes provide instructions: 1. Set values ≥ 100 as 1, click Add; 2. Set all other values as 0, click Add; 3. You should see the recoding criteria here; 4. Click OK.

1. Set values ≥ 100 as 1, click Add

2. Set all other values as 0, click Add

3. You should see the recoding criteria here

4. Click OK

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

MAKE GROUPED BAR PLOT -- I

The image shows the Plot Builder interface with the following configuration:

- Template:** grouped bar
- Variable:** Potential
- Group:** Radon_bin
- Buttons:** OK, Cancel, Run, Reset, Cancel

1. Go to Plot Builder, click on tab Templates

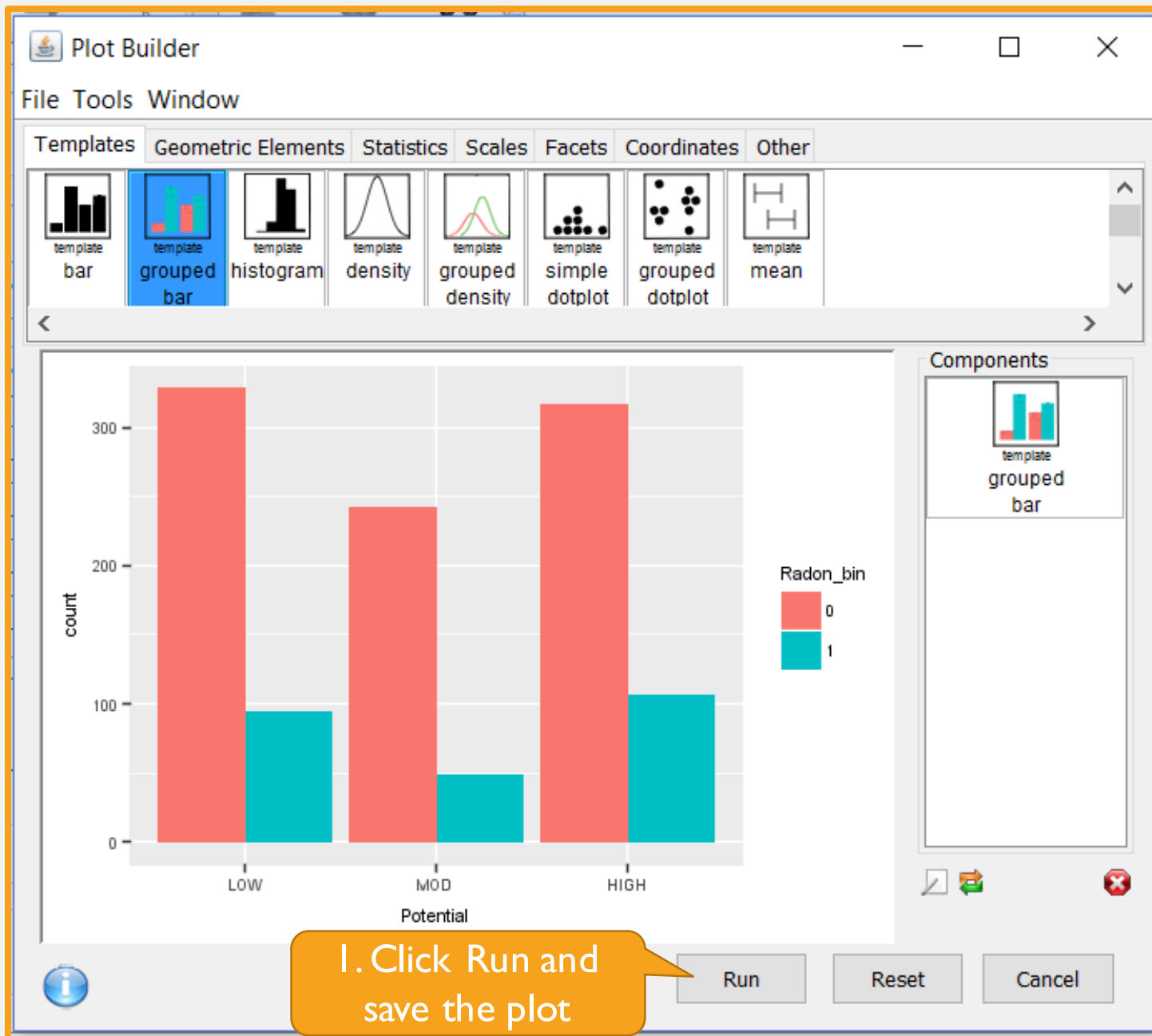
2. Double click "grouped bar"

3. Put in the variable for x axis

4. Put in the binary radon variable (it has to be factor)

5. Click OK

MAKE GROUPED BAR PLOT -- 2



CROSS-TABULATION AND CHI-SQUARED TESTS -- I

The screenshot shows the R Studio interface with the 'Contingency Tables' dialog box open. The dialog box is titled 'Contingency Tables' and has a close button (X) in the top right corner. On the left, there is a list of variables from a dataset named 'dat'. The variables are: UID, Community, MainRadon, Location, Foundation, Soil, LowerWindow, Separation, Heating, AC, HomeAge1990, RockType, MainRadon.tr, Radon_subLOD, Radon_subLODhalf, Radon_subLODhalf.tr, Radon_subLODroot2, and Radon_subLODroot2.tr. In the center, there are three blue arrow buttons pointing right, indicating the flow of variables from the list to the 'Row' and 'Column' fields. The 'Row' field contains 'Radon_bin' and the 'Column' field contains 'Potential'. To the right of these fields are three buttons: 'Cells', 'Statistics', and 'Results'. At the bottom of the dialog box, there are three buttons: 'Run', 'Reset', and 'Cancel'. The 'Run' button is highlighted with an orange callout box. There are also two orange callout boxes: one pointing to the 'Analysis > Contingency Tables' menu path and another pointing to the 'Row' and 'Column' fields.

1. Go to Analysis > Contingency Tables

2. Put in the variables for the row and columns of the cross-tabulation

3. Click Run

CROSS-TABULATION AND CHI-SQUARED TESTS -- 2

1. You should see this table in the Console

		Potential			
Radon_bin		LOW	MOD	HIGH	Row Total
0	Count	328	242	316	886
	Row %	37.020%	27.314%	35.666%	78.131%
	Column %	77.725%	83.448%	74.882%	
1	Count	94	48	106	248
	Row %	37.903%	19.355%	42.742%	21.869%
	Column %	22.275%	16.552%	25.118%	
	Column Total	422	290	422	1134
	Column %	37.213%	25.573%	37.213%	

Large Sample
Test Statistic DF p-value
Chi Squared 7.447 2 0.024

2. Also the result of the Chi-Squared test

LOGISTIC REGRESSION -- I

1. Go to Analysis
> Logistic Model

2. Select the binary
radon variable as
outcome

3. Put in any
independent
variable(s)

4. Click Continue here
and in the next page

The screenshot shows the 'Logistic Regression Model' dialog box in R Studio. The 'Outcome' field is set to 'Radon_bin'. The 'Split' field is set to 'Radon_bin==1'. The 'As Numeric' section is empty. The 'As Factor' section has 'Potential' selected. The 'Weights' and 'Subset' fields are empty. The 'Continue' button is highlighted.

dat

Filter:

UID
Community
MainRadon
Location
Foundation
Soil
LowerWindow
Separation
Heating
AC
HomeAge1990
RockType
MainRadon.tr
Radon_subLOD
Radon_subLODhalf
Radon_subLODhalf.tr
Radon_subLODroot2
Radon_subLODroot2.tr

Outcome
Radon_bin

Split Radon_bin==1

As Numeric

As Factor
Potential

Weights

Subset

Continue Reset Cancel

LOGISTIC REGRESSION -- 2

The screenshot shows the 'Logistic Model Explorer' window. The title bar indicates the model is 'Radon_bin==1' ~ Potential'. The 'General' tab is selected, and the 'Preview' section displays the following R console output:

```
Call:
glm(formula = Radon_bin == "1" ~ Potential, family = binomial(),
     data = .gui.working.env$dat, na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7606  -0.7606  -0.7099  -0.6016   1.8967

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2497    0.1170 -10.682  <2e-16 ***
PotentialMOD  -0.3680    0.1966  -1.872   0.0612 .
PotentialHIGH  0.1574    0.1621   0.971   0.3316
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1191.3  on 1133  degrees of freedom
Residual deviance: 1183.6  on 1131  degrees of freedom
AIC: 1189.6
```

On the right side of the preview window, there are several buttons: 'Options', 'Post Hoc', 'Tests', 'Plots', 'Means', 'Export', and 'Update Model'. At the bottom of the window, there are 'Run', 'Reset', and 'Cancel' buttons. An orange callout bubble points to the 'Run' button with the text: '1. Preview the model output and click Run to see it in Console'.