Module 2 - Solutions

Sam Dumble - Research Methods Support

01/02/2021

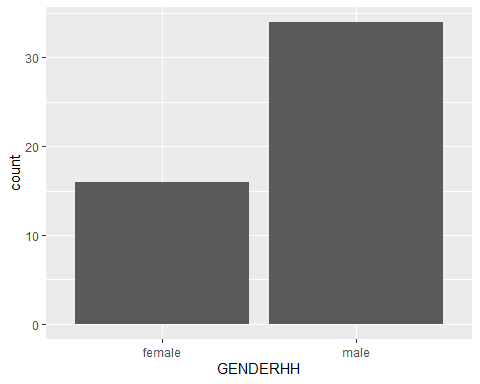
## Solutions

All of the exercises are using the BeanSurvey data we have been working with so far - remember you can also go back to look at the description of the data and the variables.

**Exercise 1. Replace each of the instances of “ZZZ” from the code below to produce a bar chart showing the frequencies of the household head gender variable GENDERHH**

ggplot(data=ZZZ, aes(x=ZZZ)) +   
 geom\_ZZZ()

ggplot(data=BeanSurvey, aes(x=GENDERHH)) +   
 geom\_bar()



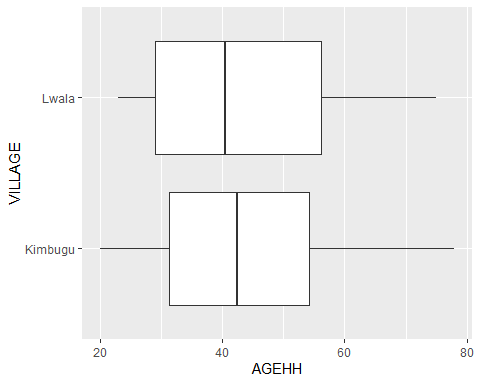
To solve the question we need to replace three ZZZ - first with the name of the data BeanSurvey and then with the name of the variable within the mapping GENDERHH. Be very careful to make sure you spell these exactly correct with upper and lower case.

We also need to set the correct geom\_ - in this case to show a bar chart of frequencies we use geom\_bar

**Exercise 2. Identify and fix the error(s) in this code to produce a boxplot of age of household head by village**

ggplot2(data = BEANSURVEY, aes(x = AgeHH, y = Village)   
 geom\_box()

ggplot(data = BeanSurvey, aes(x = AGEHH, y = VILLAGE)) +  
 geom\_boxplot()



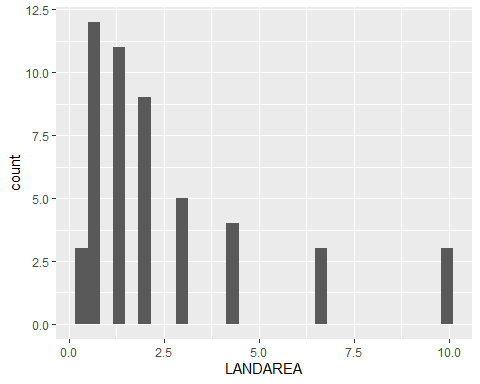
There are lots of errors!: \* There is no + at the end of the first line \* In the second line there are two brackets opened but only one is closed \* The function is ggplot not ggplot2 \* The data has been entered with the incorrect case - BeanSurvey not BeanSurvey \* The variables have been entered with incorrect cases. AGEHH not AgeHH; VILLAGE not Village \* The correct function is geom\_boxplot()

**Exercise 3. Produce a histogram of land area. Set the bins so that each bin covers a range of 1 acre**

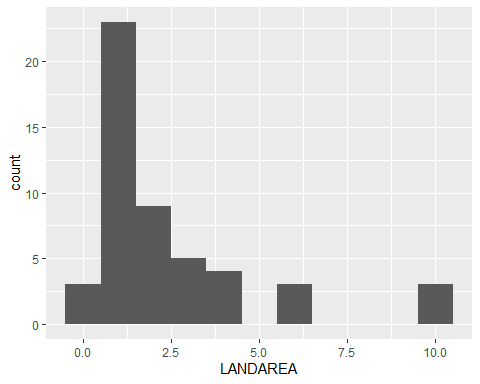
We first need to convert the question into the different ggplot elements: data-> BeanSurvey aesthetics->x->LANDAREA geom->histogram

ggplot(data=BeanSurvey,aes(x=LANDAREA))+  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 Once we have this we can think about modifying to only cover 1 acre per bin. We need to find the option binwidth from within the geom\_histogram function

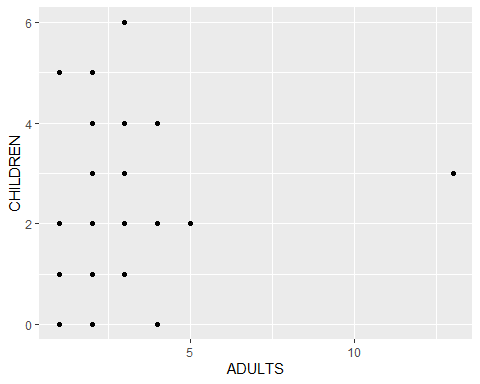
ggplot(data=BeanSurvey,aes(x=LANDAREA))+  
 geom\_histogram(binwidth = 1)



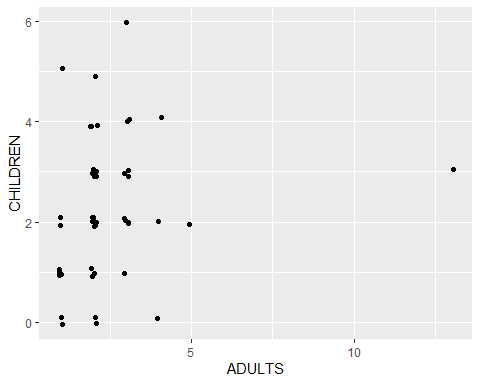
**Exercise 4. Make a scatter plot showing the number of adults on the x axis against the number of children on the y axis. Can you see any limitations to this plot?**

data-> BeanSurvey aesthetics->x->CHILDREN aesthetics->y->ADULTS geom->point

ggplot(data=BeanSurvey,aes(x=ADULTS,y=CHILDREN))+  
 geom\_point()

 A limitation here is that we have lots of overlapping observations. e.g. lots of people in our data have 2 adults and 3 children. But we only see one point. This could be a case where we might want to do a jitter plot instead.

ggplot(data=BeanSurvey,aes(x=ADULTS,y=CHILDREN))+  
 geom\_jitter(width=0.1,height=0.1)

 The width and height control how much away from the centre the points are allowed to jitter