Chapter 8 - Graphics

Exercises

Brian Fogarty

15 May 2018

Contents

EXERCISE I - BAR PLOTS

Using the 2012 survey about smoking and drug use amongst English pupils (`2012smokedrugs.dta`), create the following bar plots using `ggplot2` code.

1. Create a bar plot with grey bars using the recoded version of `lifewant` from Chapter 5, which we called `lifewant2`. Discuss what it shows you.
2. Create the same bar plot but now with red bars.
3. Create a bar plot with `lifewant2` and `free`. You should label the values of `free` as `0='0. Not Free Lunch';1='1. Free Lunch'`. Discuss what it shows you.
4. Create the same bar plot but now with grey shade bars.

5. Create the same bar plot but now using dodging.

ANSWERS FOR EXERCISE I

```
setwd("C:/QSSD/Exercises/Chapter 8 - Exercises")
getwd()
```

```
[1] "C:/QSSD/Exercises/Chapter 8 - Exercises"
```

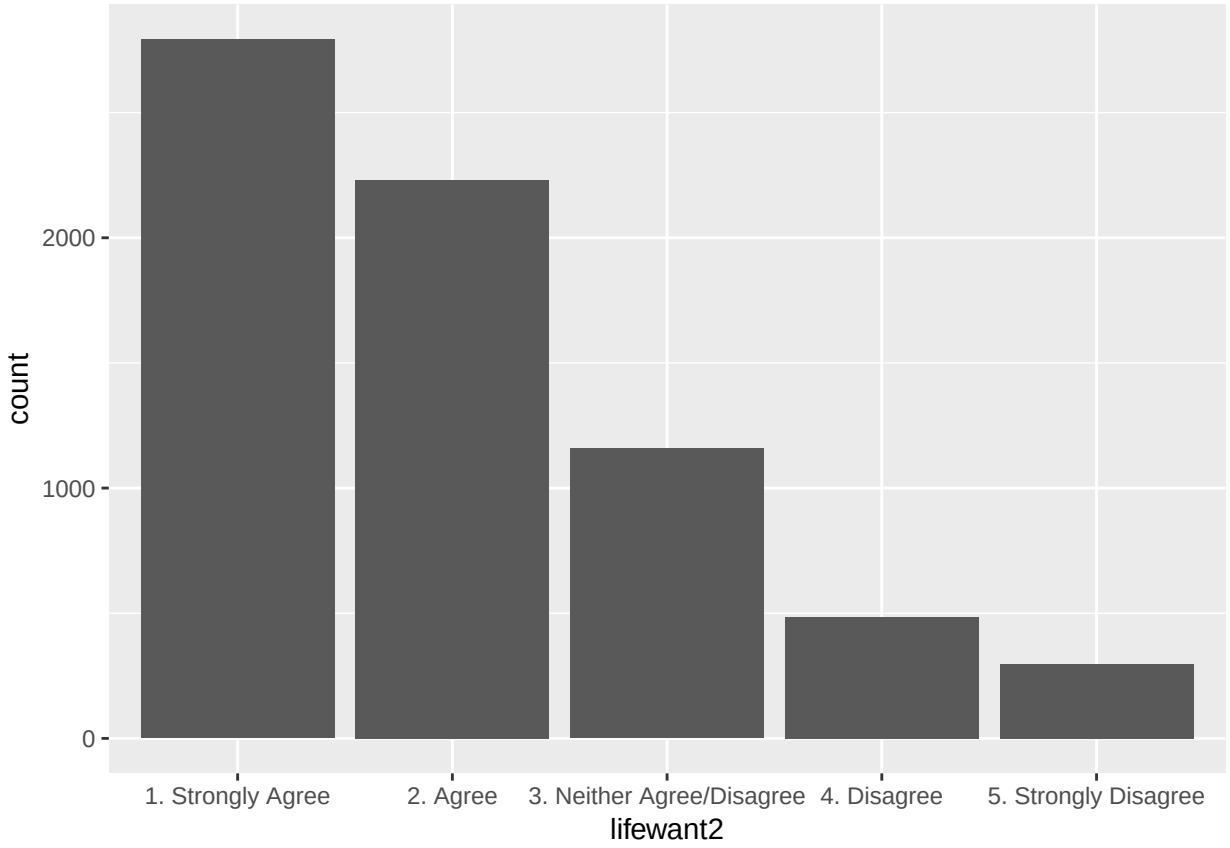
```
library(foreign)
drugs <- read.dta("2012smokedrugs.dta", convert.factors=FALSE)
```

Question 1.1

```
library(car)
```

```
Warning: package 'car' was built under R version 3.4.3
```

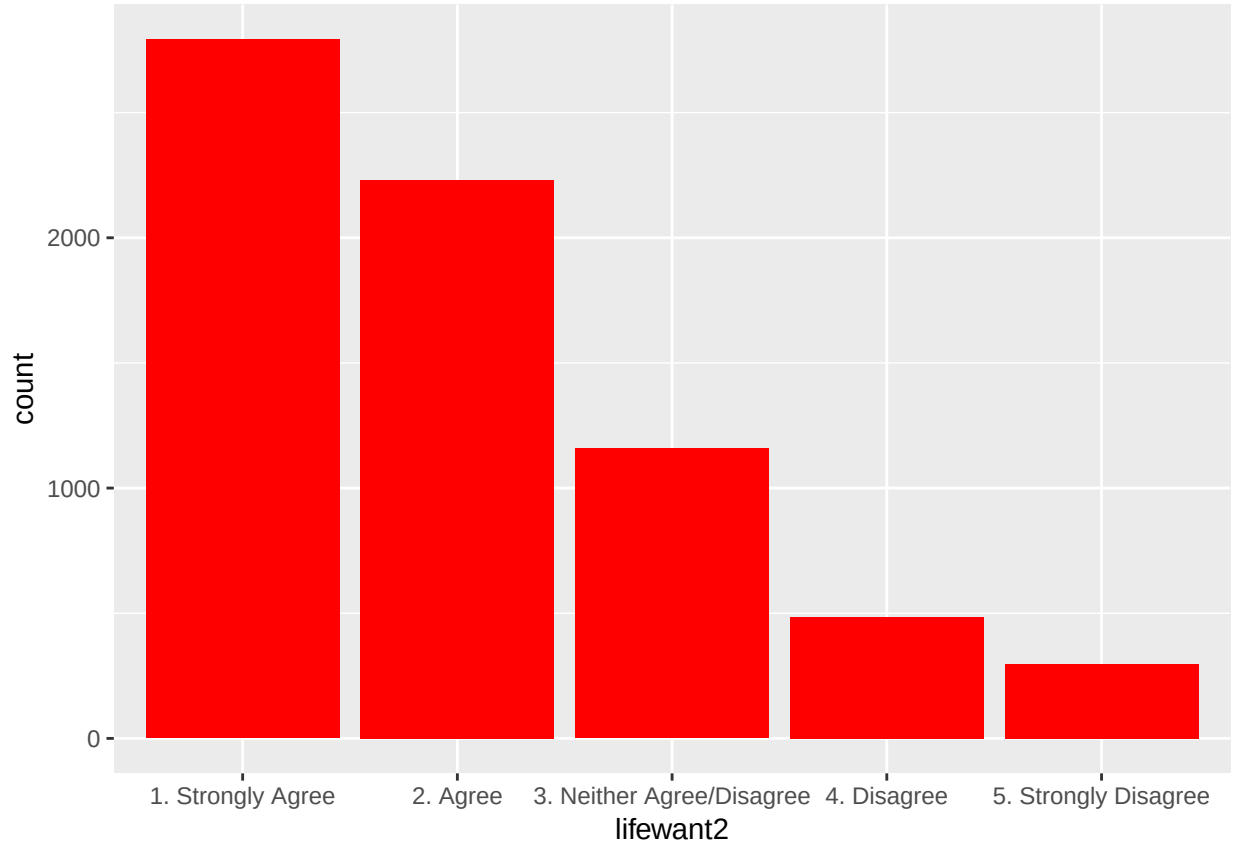
```
drugs$lifewant2 <- recode(drugs$lifewant, "1='1. Strongly Agree';2='2. Agree';
                          3='3. Neither Agree/Disagree';
                          4='4. Disagree';5='5. Strongly Disagree'")

drugs1 <- subset(drugs, !is.na(lifewant2))

library(ggplot2)
x11()
ggplot(data = drugs1) +
  geom_bar(mapping = aes(lifewant2))
```



The bar plot shows that the modal response is that pupils "strongly agree" that they have the life they want. Meanwhile, the smallest category are pupils responding "strongly disagree" that they have the life they want.

Question 1.2

```
x11()
ggplot(data = drugs1) +
  geom_bar(mapping = aes(lifewant2), fill="red")
```

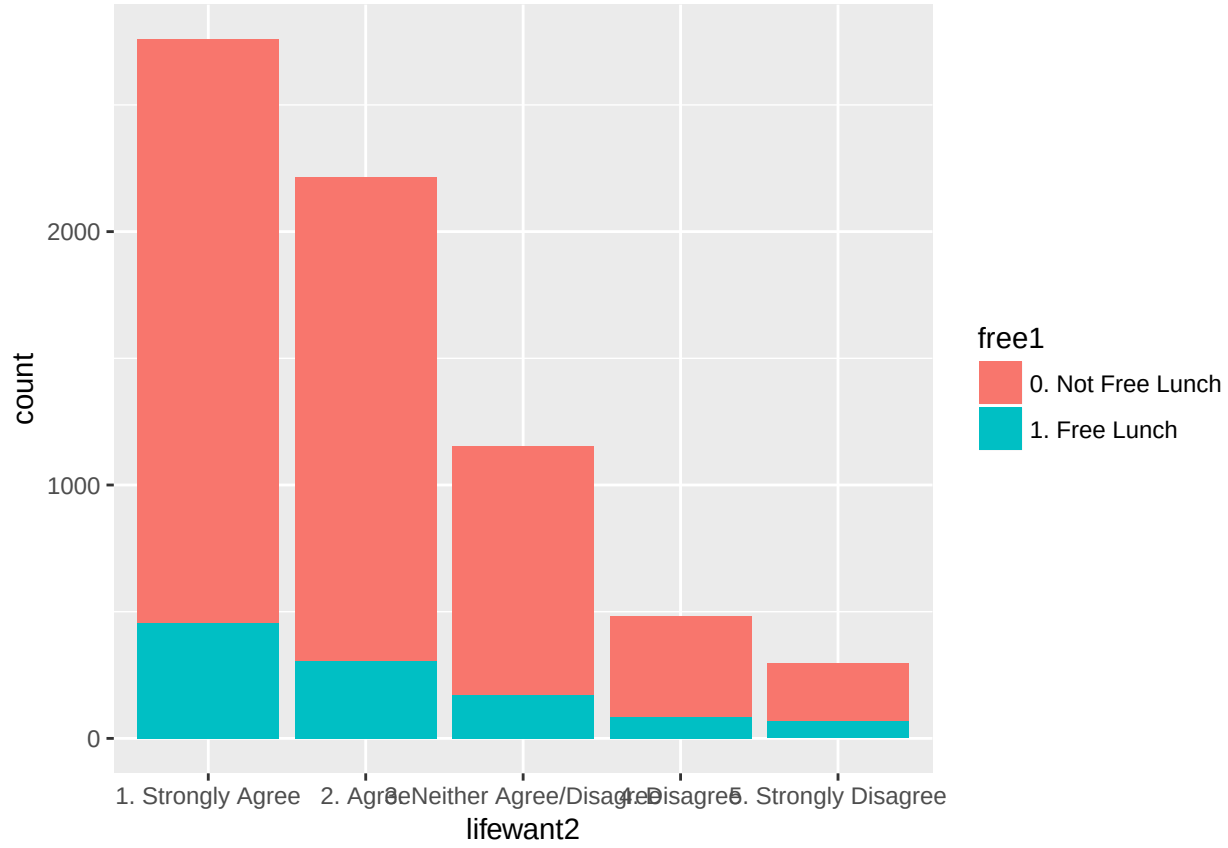


Question 1.3

```
drugs$free1 <- recode(drugs$free, "0='0. Not Free Lunch';1='1. Free Lunch'")
table(drugs$free1)
```

```
0. Not Free Lunch     1. Free Lunch
             6162              1194
```

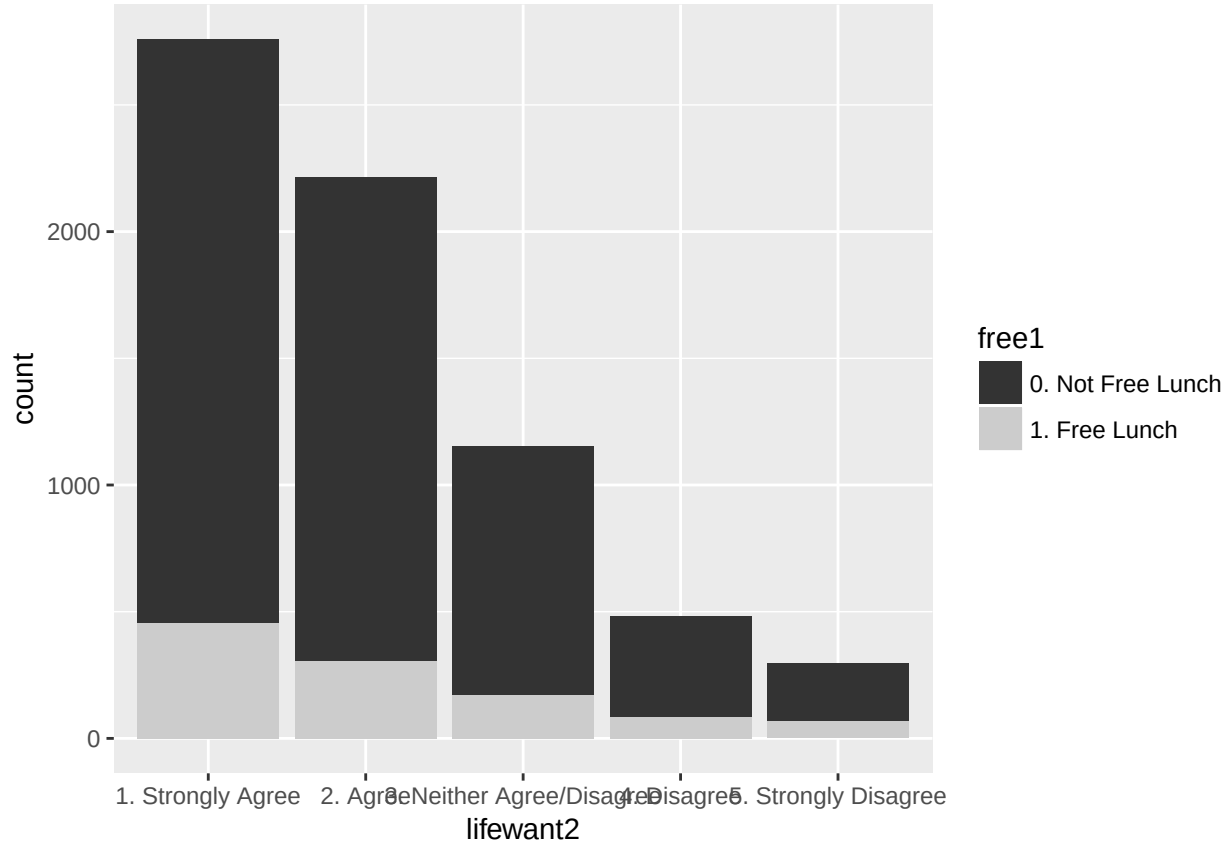
```
drugs1 <- subset(drugs, !is.na(lifewant2) & !is.na(free1))

x11()
ggplot(data = drugs1) +
  geom_bar(mapping = aes(lifewant2, fill=free1))
```



The bar plot shows that whether a student receives a free lunch or not makes up roughly the same portion of responses for each category.

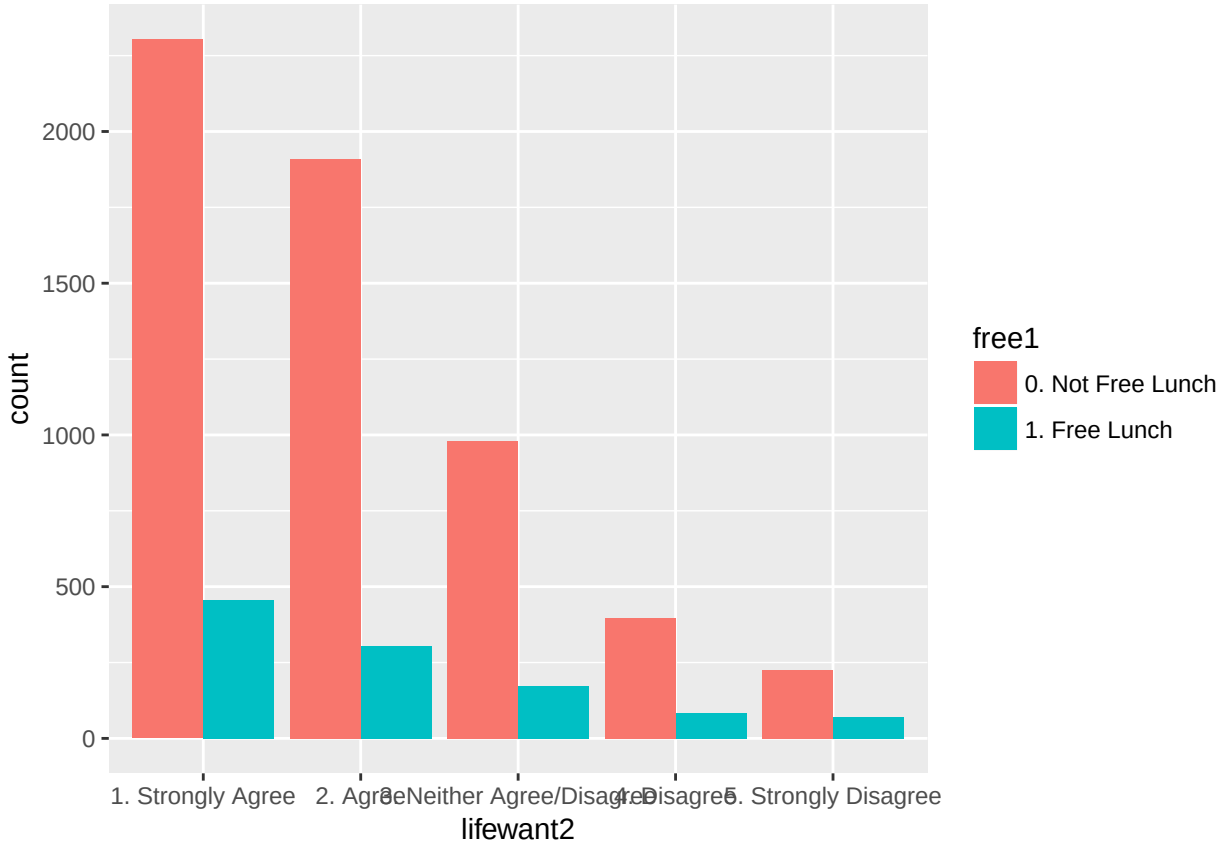
Question 1.4

```
x11()
ggplot(data = drugs1) +
  geom_bar(mapping = aes(lifewant2, fill=free1)) +
  scale_fill_grey()
```



Question 1.5

```
x11()
ggplot(data = drugs1) +
  geom_bar(mapping = aes(lifewant2, fill=free1), position="dodge")
```



EXERCISE II - HISTOGRAMS

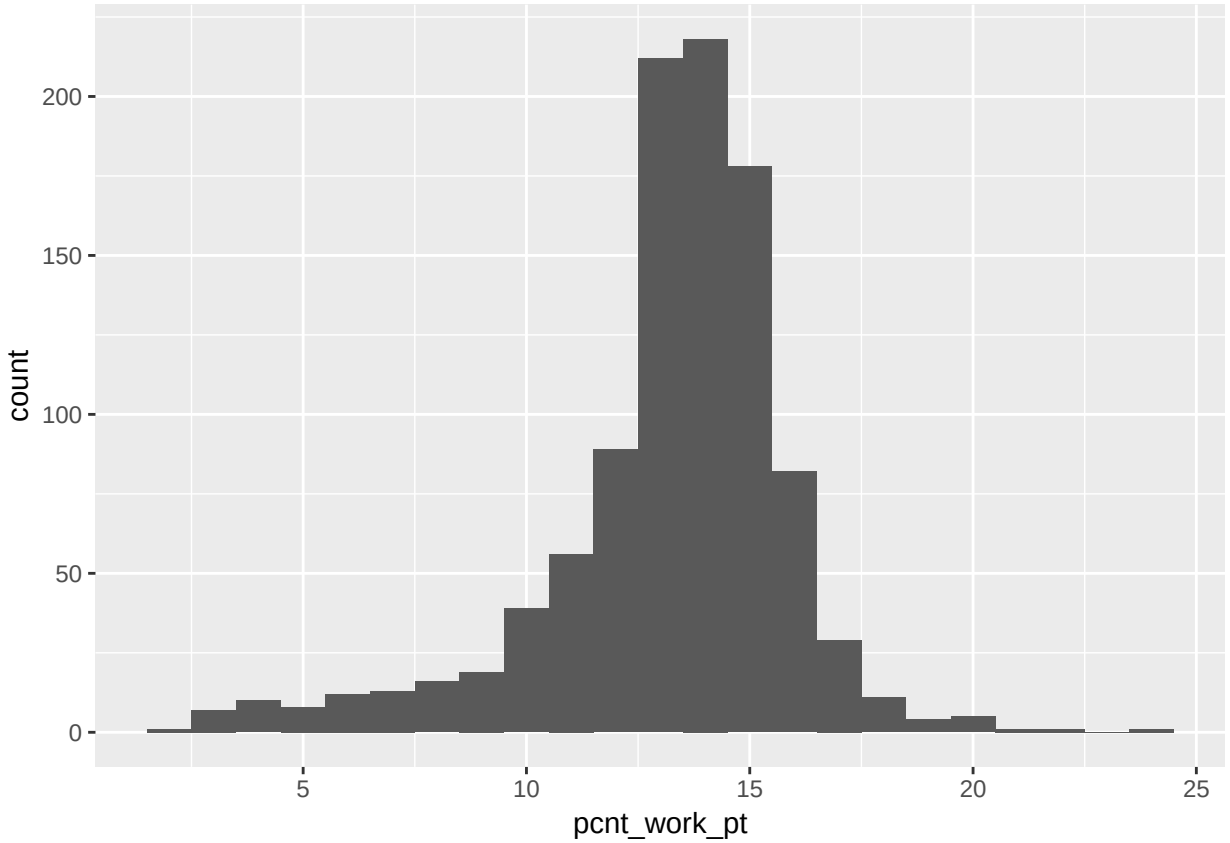
Using the 2011 Scottish postcodes data (`depdata.dta`), create the following historgrams using `ggplot2` code.

1. Create a histogram with using the variable `pcnt_work_pt`, which is the percentage of people working part-time, and set the binwidth to 1. Discuss what it shows you.
2. Create the same histogram but now with red bars.
3. Create a histogram with `pcnt_work_pt` and `urban` as the second variable. Discuss what it shows you.

ANSWERS FOR EXERCISE II

Question 2.1

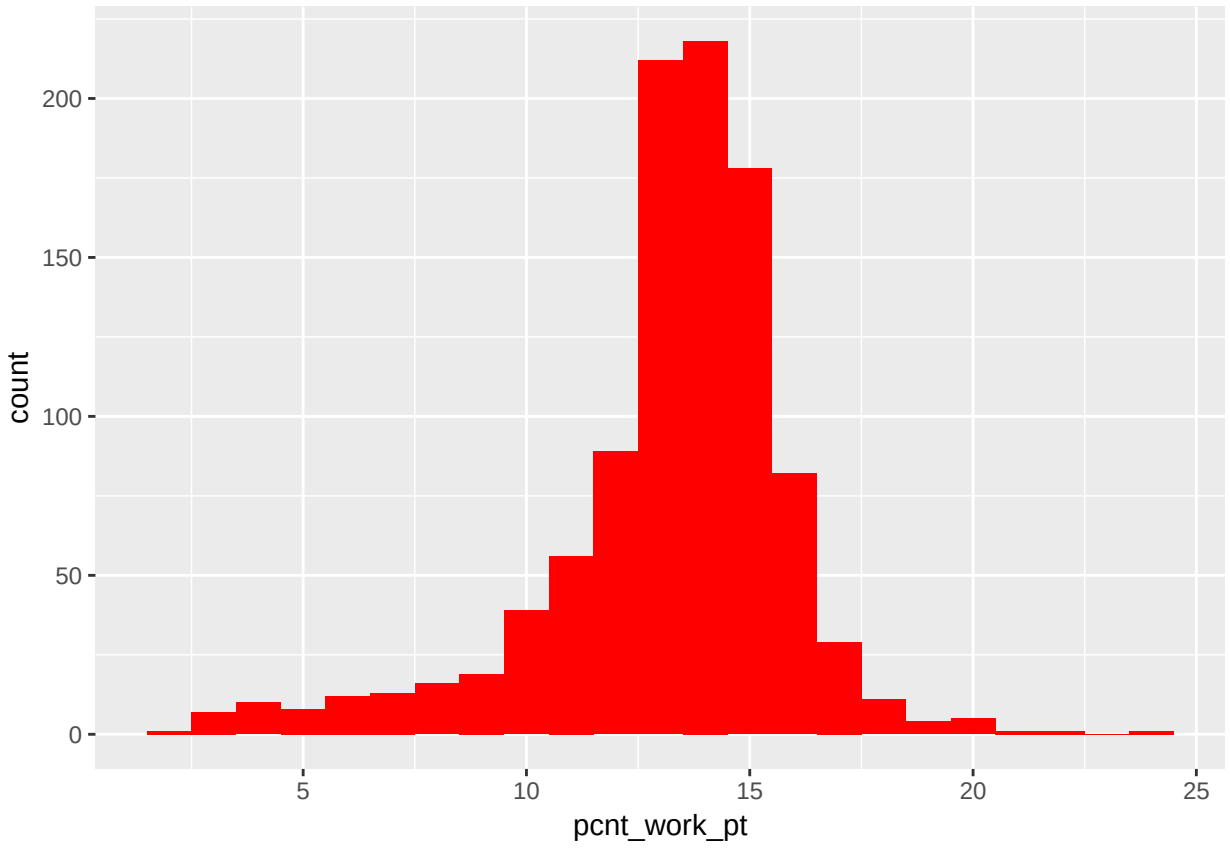
```
depdata <- read.csv("depdata.csv", na="NA")

x11()
ggplot(data=depdata) +
  geom_histogram(mapping = aes(pcnt_work_pt), binwidth=1)
```



The histogram shows that most postcodes are clustered around 12-16% percentage of individuals working part-time.

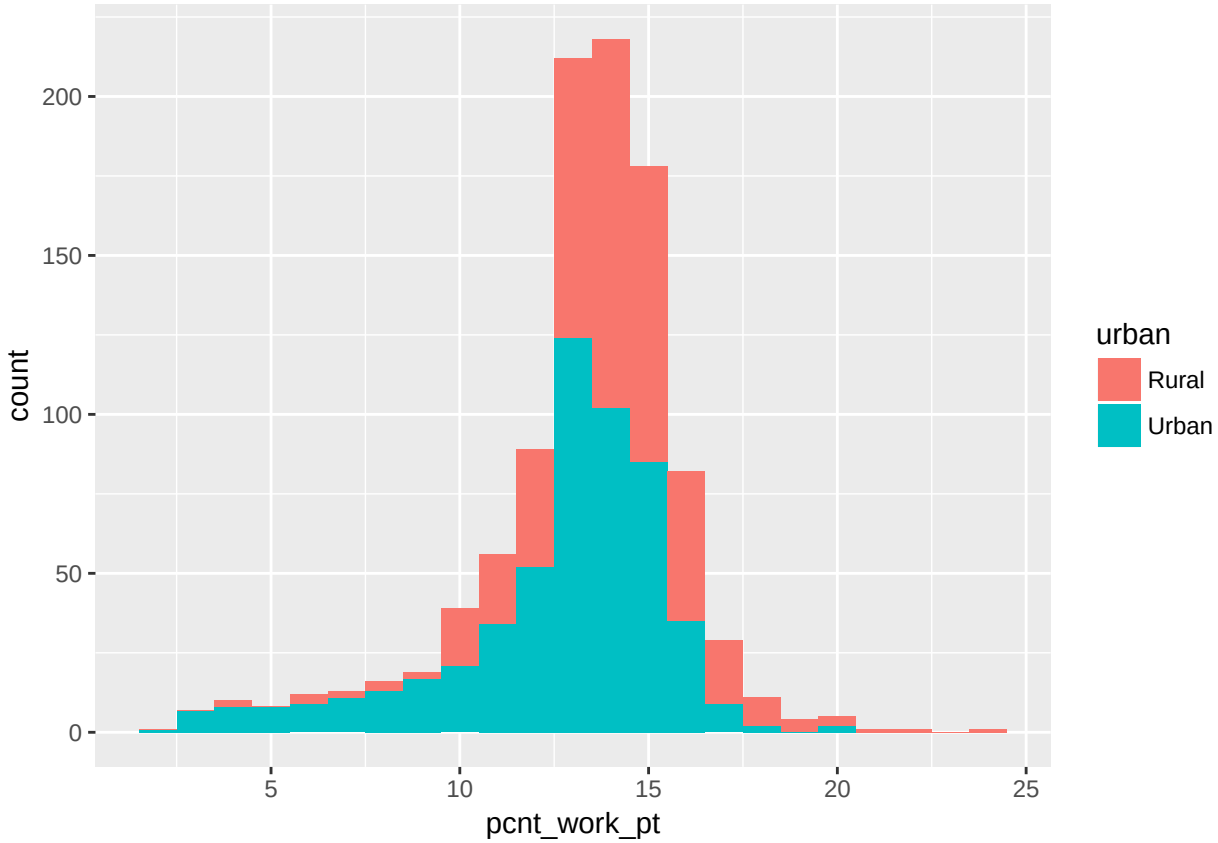
Question 2.2

```
x11()
ggplot(data=depdata) +
  geom_histogram(mapping = aes(pcnt_work_pt), binwidth=1, fill="red")
```



Question 2.3

```
x11()
ggplot(data=depdata) +
  geom_histogram(mapping = aes(pcnt_work_pt, fill=urban),
                 binwidth=1)
```



The histogram shows, though not dramatically different, there are more urban postcodes with a small percentage of people working part-time than rural postcodes. Further, there are more rural postcodes with a high percentage of people working part-time than urban postcodes.

EXERCISE III - SCATTERPLOTS

Using the 2011 Scottish postcodes data (`depdata.dta`), create the following scatterplots using `ggplot2` code.

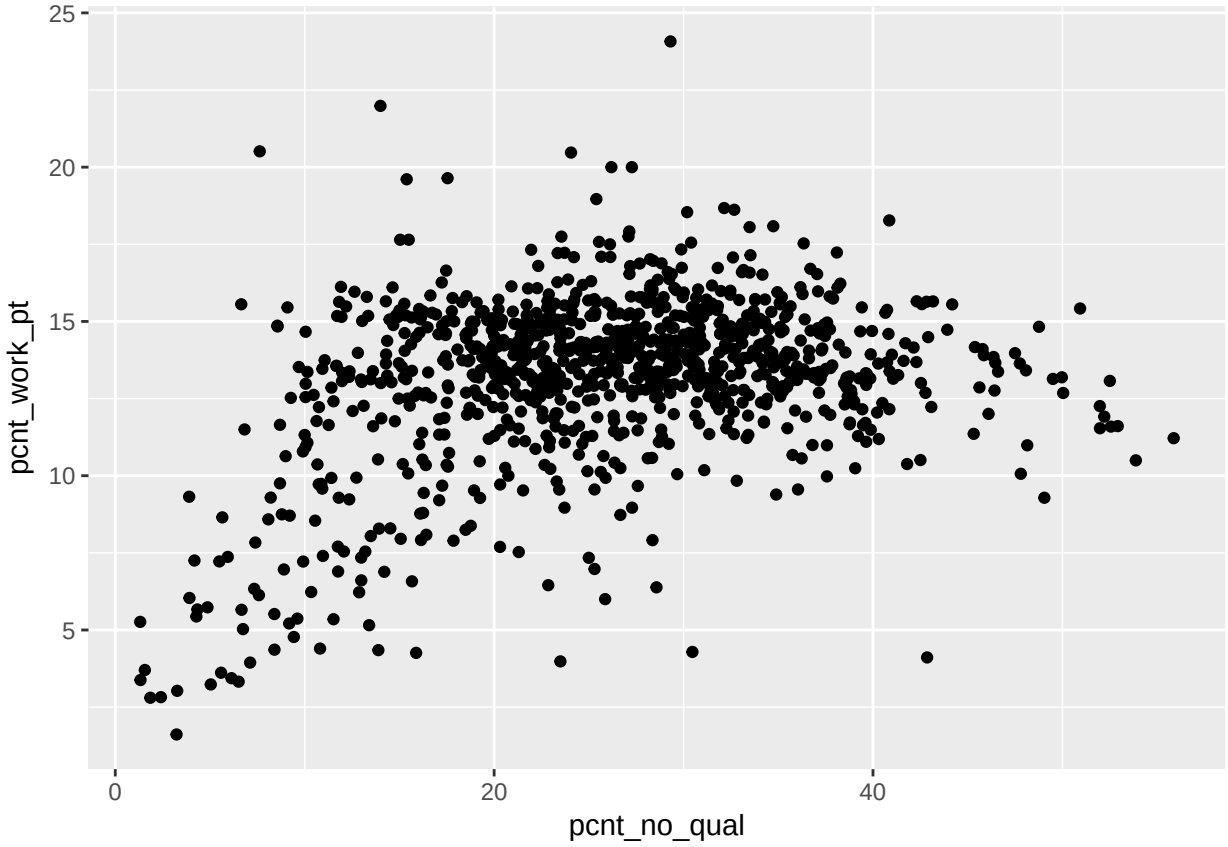
1. Create a scatterplot with black points with `pcnt_no_qual` on the x-axis and `pcnt_work_pt` on the y-axis. Discuss what it shows you.
2. Create the same scatterplot, but with red instead of black data points.
3. Create the same scatterplot using black points and `alpha=1/10` to illustrate points that overlay one another. Discuss what you find.
4. Create a scatterplot with `pcnt_no_qual` on the x-axis, `pcnt_work_pt` on the y-axis, and `urban` as the third variable. Use `scale_colour_grey()` to differentiate the urban/rural data points. Discuss what it shows you.
5. Create the same scatterplot, but with the default colours to differentiate the urban/rural data points.
6. Create the same scatterplot, but specify "darkblue" and "orange" as the colours to differentiate the urban/rural data points.
7. Create the same scatterplot, but with default symbols to differentiate the urban/rural data points.

8. Create a similar scatterplot using `urban` as the **faceting** variable. This will create two scatterplots - one for rural postcodes and one for urban postcodes.
9. Place the scatterplots from questions 4 through 7 onto one page using the `grid.arrange()` function.

ANSWERS FOR EXERCISE III

Question 3.1

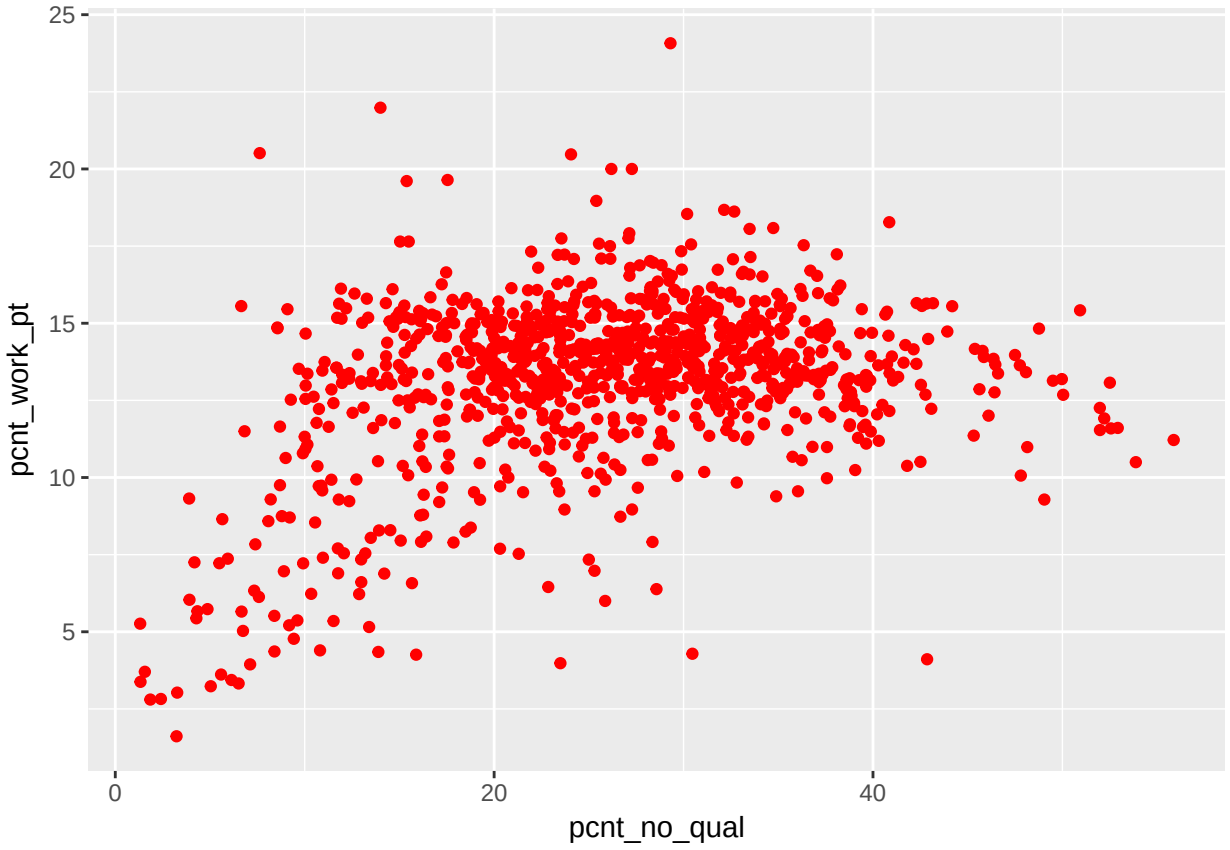
```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt))
```



The scatterplot shows that postcodes with higher percentages of people with no education qualifications also have higher percentages of people working part-time.

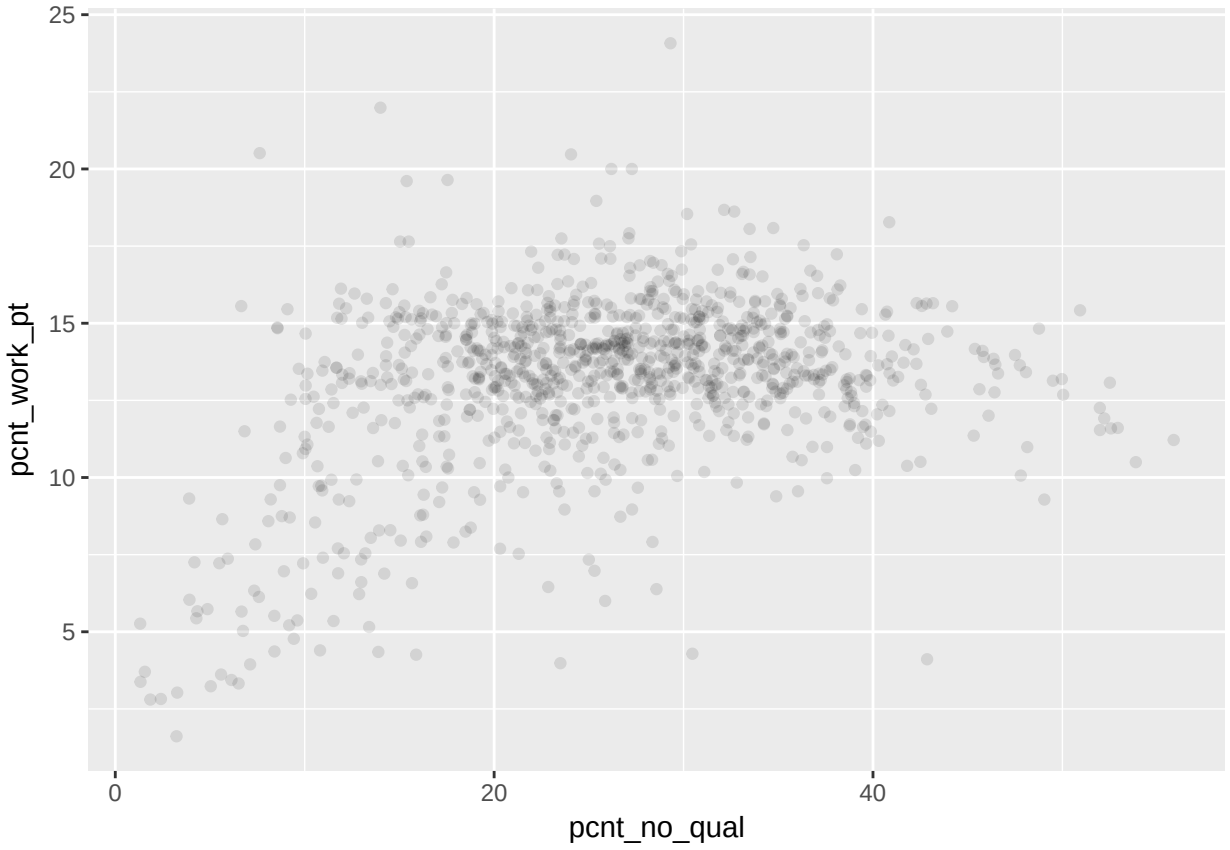
Question 3.2

```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt), colour="red")
```



Question 3.3

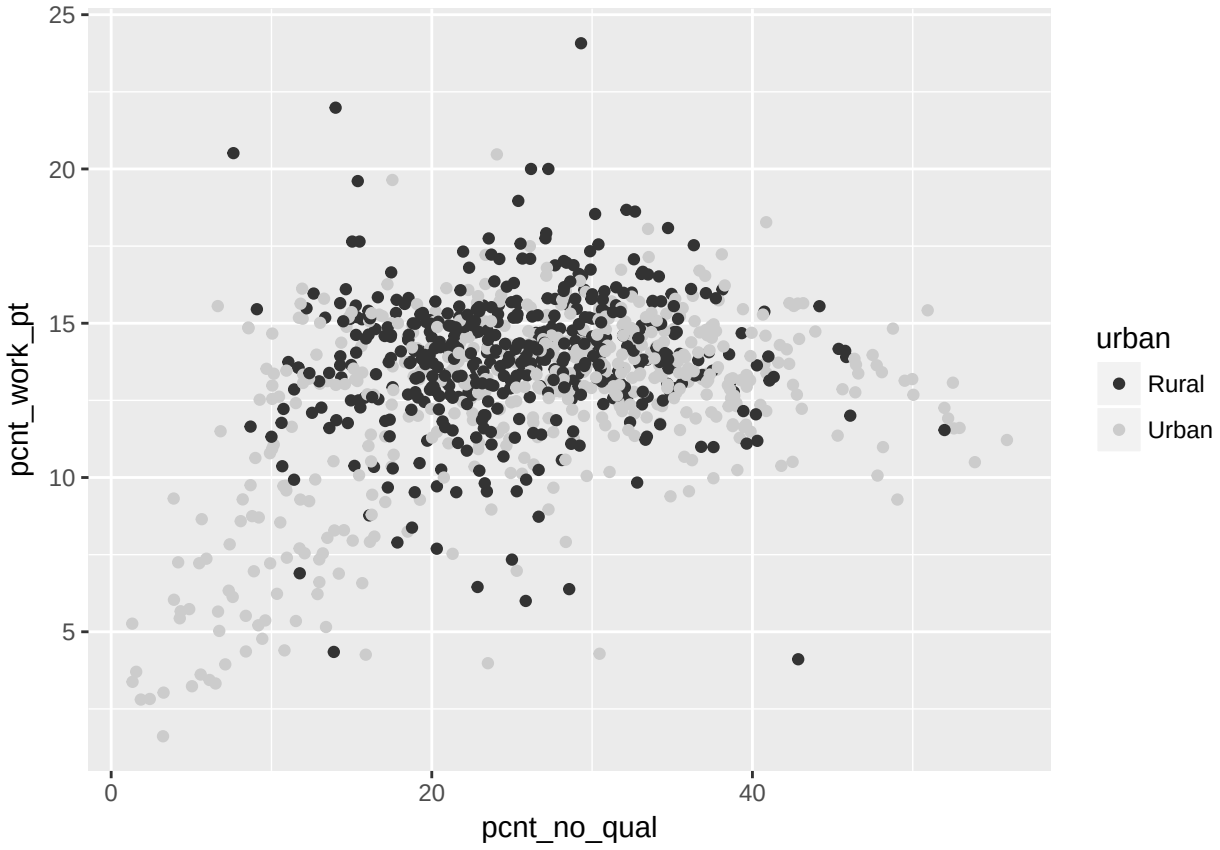
```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt), alpha=1/10)
```



This scatterplot shows a cluster of postcodes between 20-30% of people with no education qualifications and between 12-16% of people working part-time.

Question 3.4

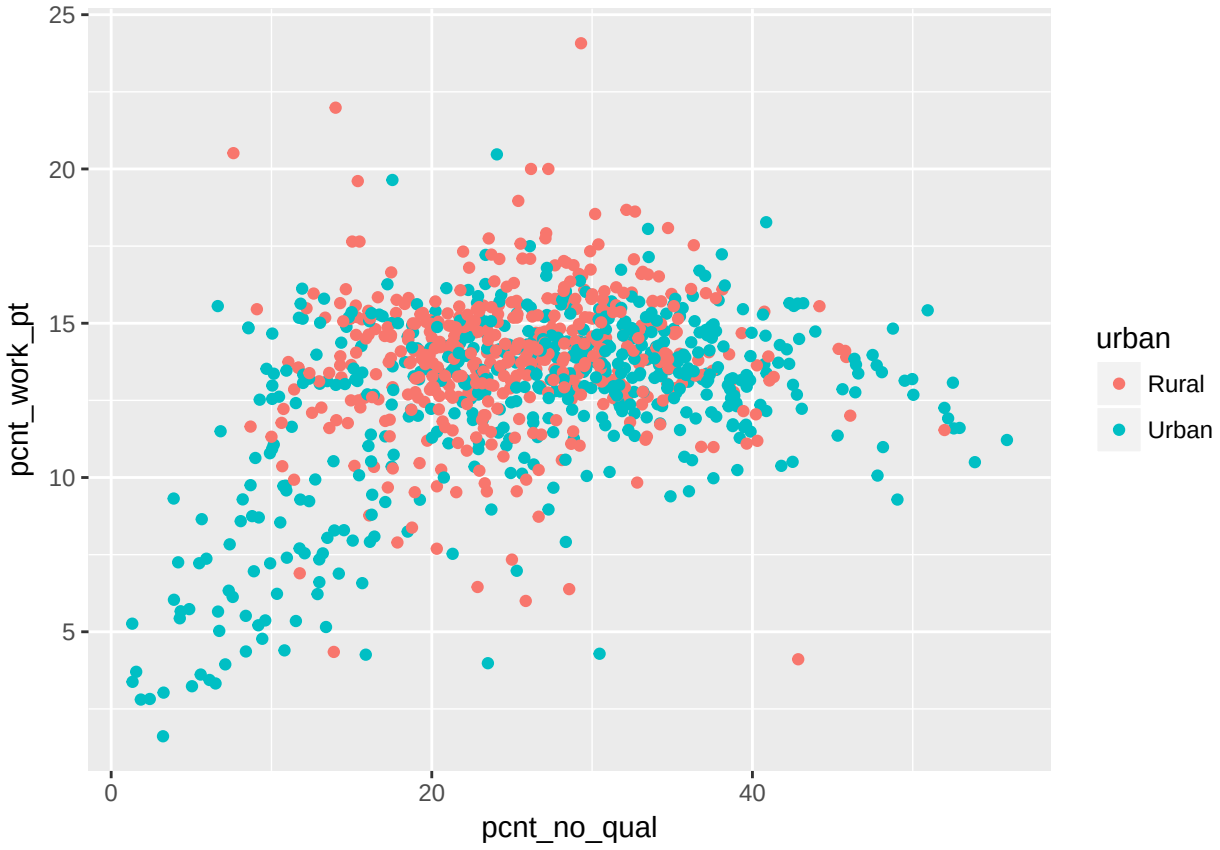
```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             colour=urban)) +
  scale_colour_grey()
```



This scatterplot shows that urban postcodes make up most of postcodes that have low percentages of people with education qualifications and working part-time.

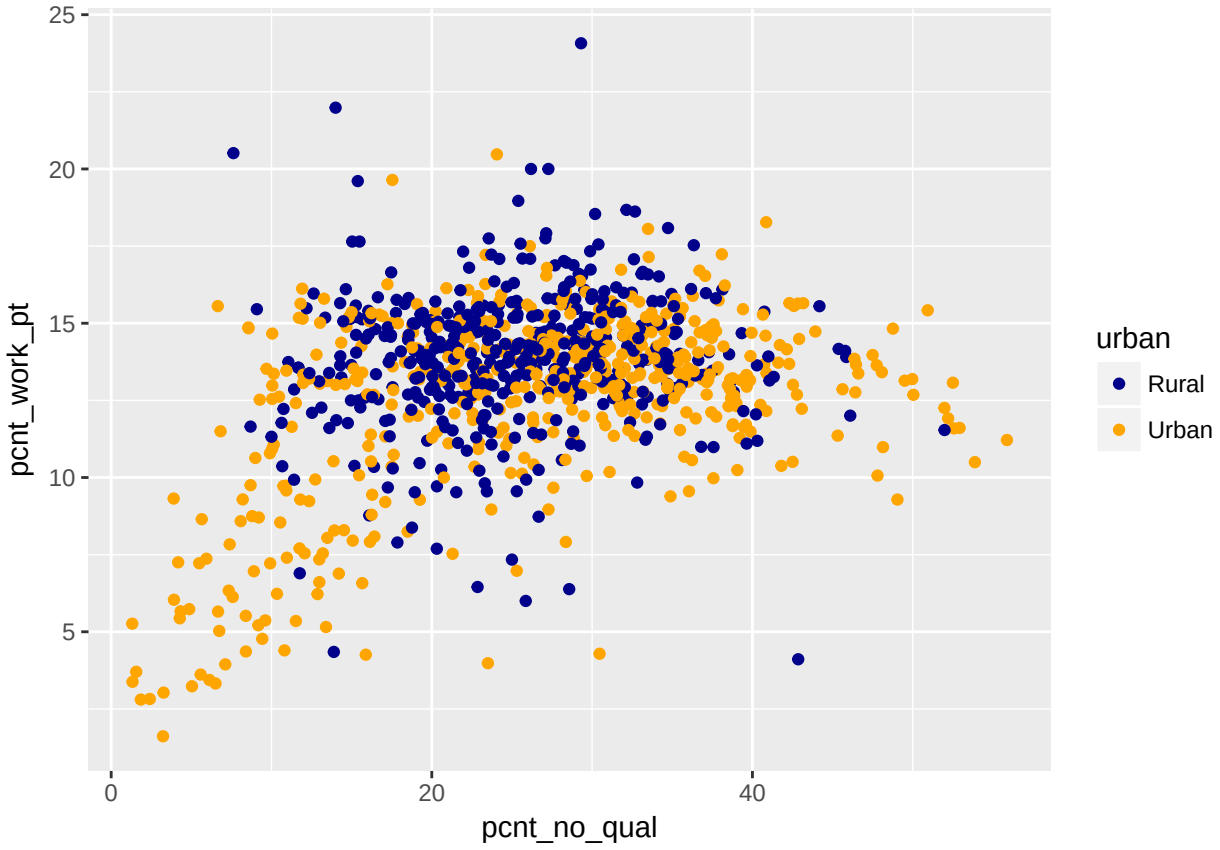
Question 3.5

```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             colour=urban))
```



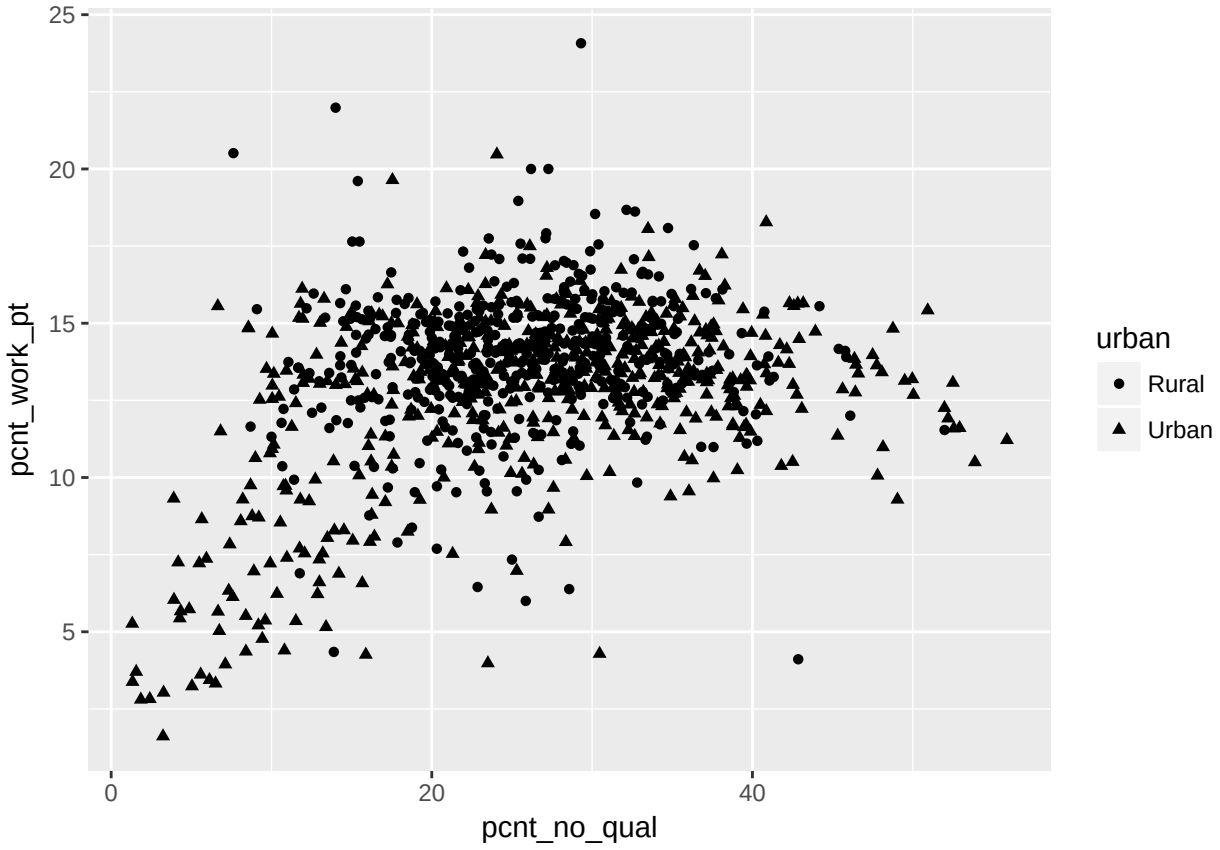
Question 3.6

```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             colour=urban)) +
  scale_colour_manual(values=c("darkblue","orange"))
```



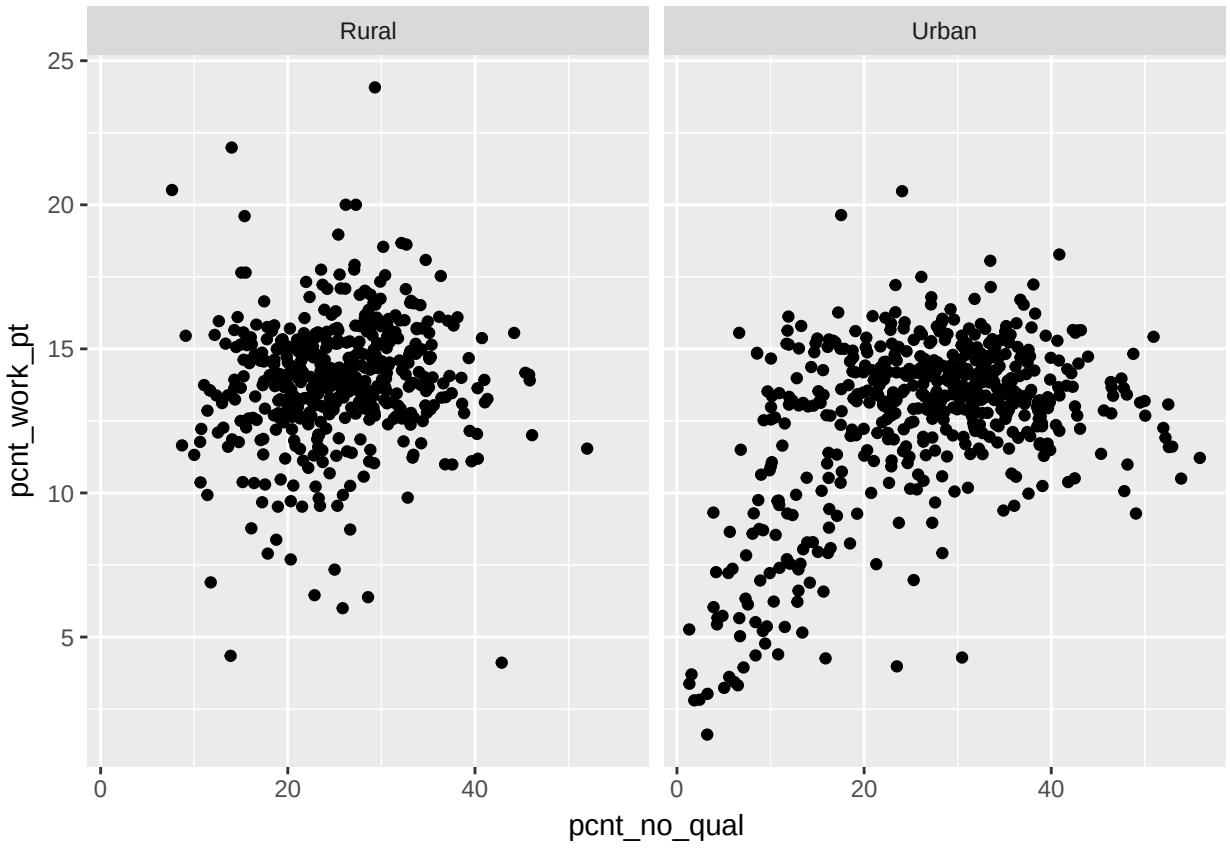
Question 3.7

```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             shape=urban))
```



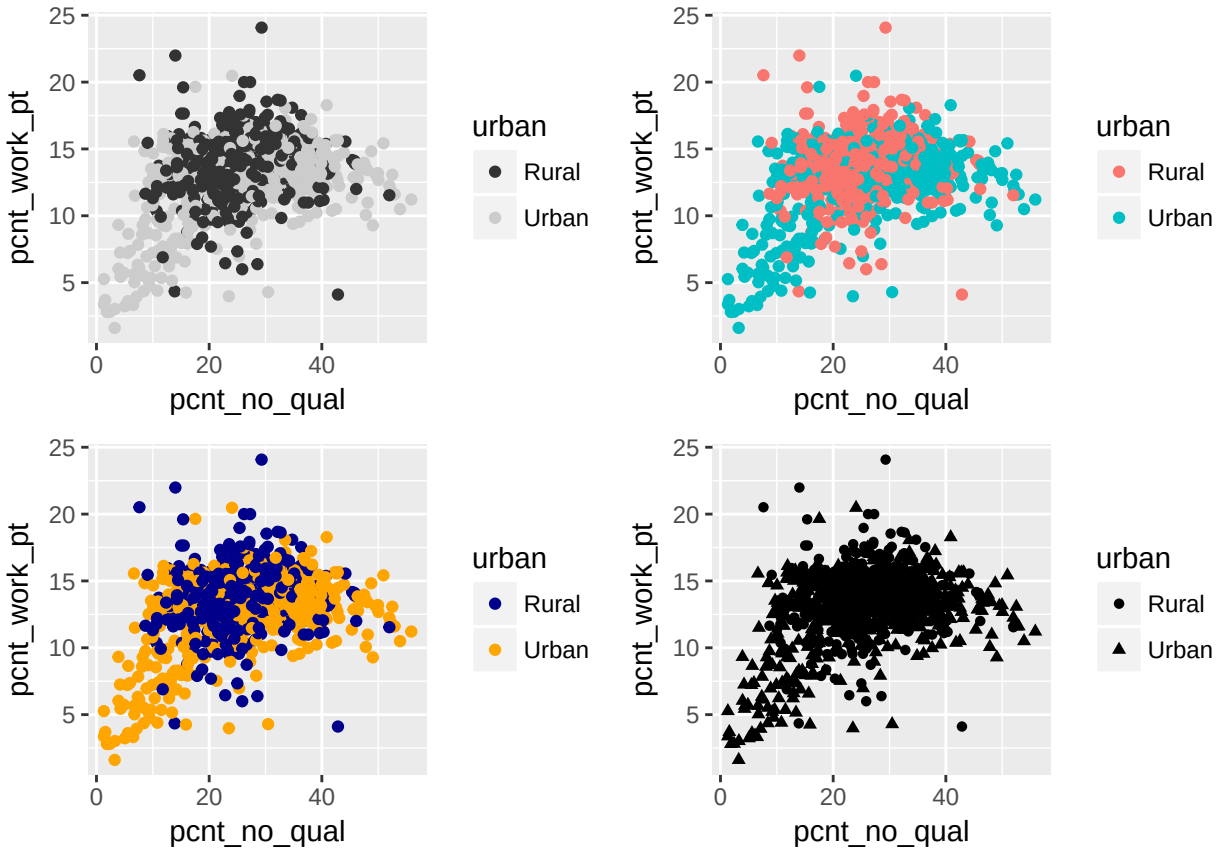
Question 3.8

```
x11()
ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt)) +
  facet_wrap(~urban)
```



Question 3.9

```
p1 <- ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             colour=urban)) +
  scale_colour_grey()

p2 <- ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             colour=urban))

p3 <- ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             colour=urban)) +
  scale_colour_manual(values=c("darkblue","orange"))

p4 <- ggplot(data=depdata) +
  geom_point(mapping=aes(x=pcnt_no_qual,y=pcnt_work_pt,
             shape=urban))

library(gridExtra)
x11()
grid.arrange(p1, p2, p3, p4, ncol=2, nrow=2)
```


pcnt_work_pt
pcnt_no_qual
urban
Rural
Urban
0
20
40
5
10
15
20
25