

Presenting Data - Quiz

Solutions

Question 1

What is the code to produce a basic pie chart in R?

- a. *p.chart()*
- b. *piechart()*
- c. *pie()*
- d. *chart()*

SOLUTION: c.

Question 2

What does the following R command do: *par(mfrow=c(1,2))*?

- a. It creates a numerical vector with the numbers 1 and 2 stored in an object called “par”
- b. It performs a parametric test between the variables 1 and 2 of a dataset called “mfrow”
- c. It displays two plots graphed in R after this command side by side
- d. It adds the numbers 1 and 2 together

SOLUTION: c.

Question 3

What kind of output would you expect from the following R command: *table(mydata[,6])*?

SOLUTION: You would expect R to produce a frequency table of the data in the 6th variable/column in your data set.

Question 4

What kind of output would you expect from the following R command: *prop.table(table(mydata[,2]))*100*?

SOLUTION: You would expect R to calculate a table with the proportions of the data in the 2nd variable/column in your data set.

Question 5

What argument do you use to adjust the length of the y-axis of a plot in R?

- a. *xlab=*
- b. *ylab=*
- c. *ylim=*
- d. *main=*

SOLUTION: c.

Question 6

What argument do you use to add a label to the y-axis of a plot in R?

- a. *xlab=*
- b. *ylab=*
- c. *ylim=*
- d. *main=*

SOLUTION: b.

Question 7

What argument do you use to add a label to the x-axis of a plot in R?

- a. *xlab=*
- b. *ylab=*
- c. *ylim=*
- d. *main=*

SOLUTION: a.

Question 8

What argument do you use to add an overall title to a plot in R?

- a. *xlab=*
- b. *ylab=*
- c. *ylim=*
- d. *main=*

SOLUTION: d.

Question 9

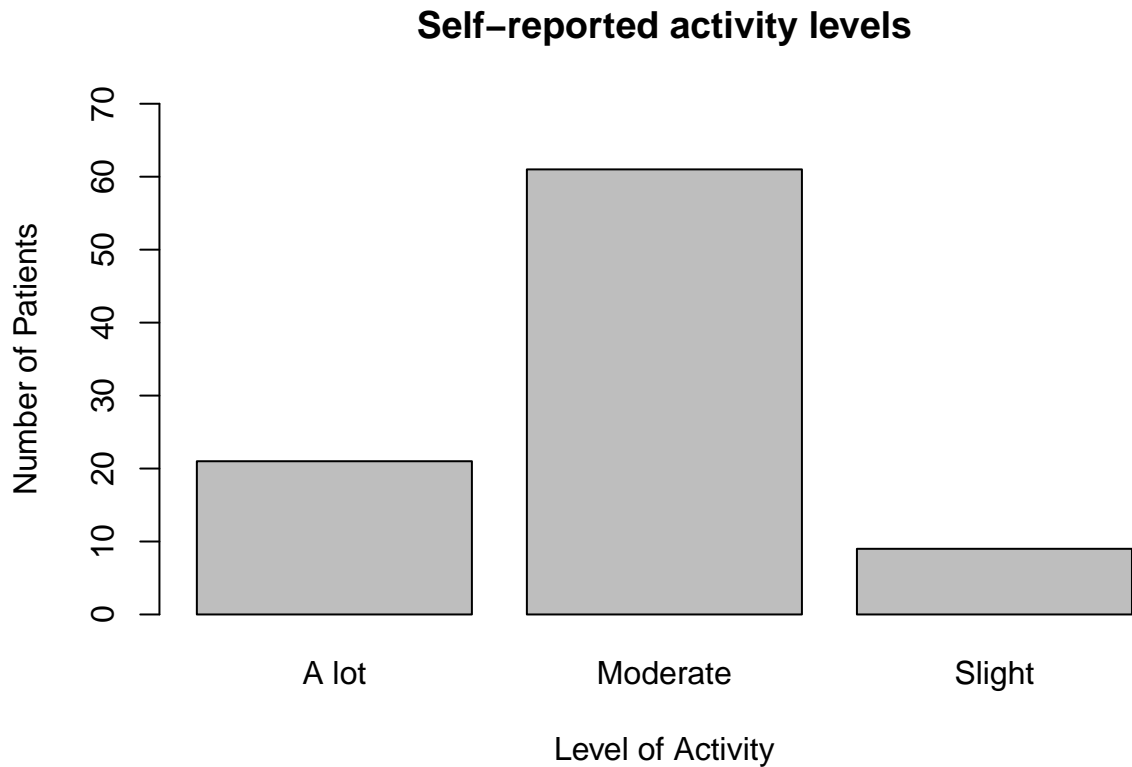
What is the code to produce a basic histogram in R?

- a. *histogram()*
- b. *histo()*
- c. *hist.plot()*
- d. *hist()*

SOLUTION: d.

Question 10

What code was used to produce the following bar plot in R?



To produce the plot, the following data were stored in an object called “levels”:

```
##  
##   A lot Moderate   Slight  
##     21      61      9
```

SOLUTION: `barplot(levels, main="Self-reported activity levels", xlab="Level of Activity", ylab="Number of Patients", names.arg=c("A lot", "Moderate", "Slight"), ylim = c(0,70))`

Question 11

What is the code to produce a basic scatter plot in R?

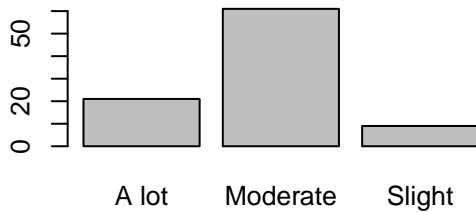
- `scatter()`
- `plot()`
- `scatterplot()`
- `sct.plot()`

SOLUTION: b.

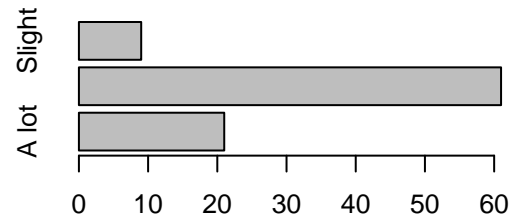
Question 12

Match each of the following four bar plot with the corresponding code segments (A, B, C and D) that were used to produce the bar plot in R:

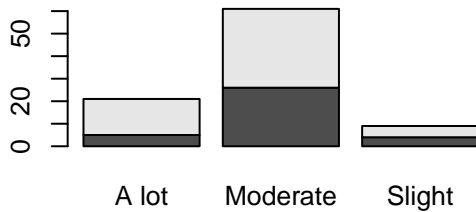
1 – Simple Bar Plot



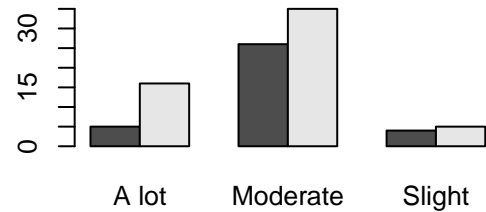
2 – Simple Horizontal Bar Plot



3 – Stacked Bar Plot



4 – Grouped Bar Plot



- A. `barplot(counts2, main="XXX")`
- B. `barplot(counts, horiz = TRUE, main="XXX")`
- C. `barplot(counts2, beside=TRUE, main="XXX")`
- D. `barplot(counts, main="XXX")`

SOLUTION:

Plot	R Code
1	D
2	B
3	A
4	C

Question 13

Which of the following can be used to visually explore the distribution/spread of data within one numerical variable?

- a. *hist()*
- b. *boxplot()*
- c. *barplot()*
- d. *pie()*

SOLUTION: a. and b.

Question 14

Which of the following R commands can be used to find measures of location?

- a. *mean()*, *median()*, *table()*
- b. *average()*, *median()*, *table()*
- c. *mean()*, *median()*, *mode()*
- d. *mean()*, *med()*, *mode()*

SOLUTION: a.

Question 15

Which of the following is NOT a way of calculating the mean in R?

- a. *sum(data\$variable)/dim(data)[1]*
- b. *avg(data\$variable)*
- c. *sum(datavariable)/length(datavariable)*
- d. *mean(data\$variable)*

SOLUTION: b.

Question 16

Which of the following are ways to calculate measures of spread in R?

- a. *sd(data\$variable)*
- b. *var(data\$variable)*
- c. *IQR(data\$variable)*
- d. All of the above

SOLUTION: d.

Question 17

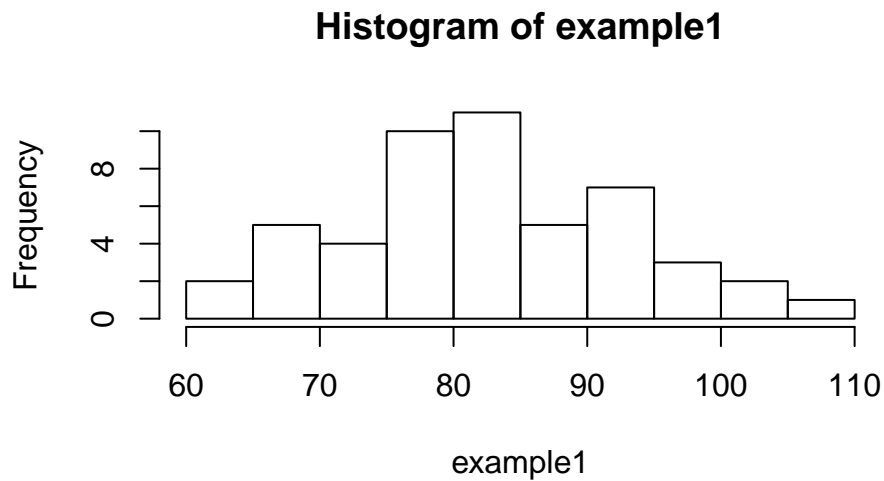
Which TWO of the following measures are highly sensitive to outliers?

- a. Mean
- b. Median
- c. Standard Deviation
- d. IQR

SOLUTION: a. and c.

Question 18

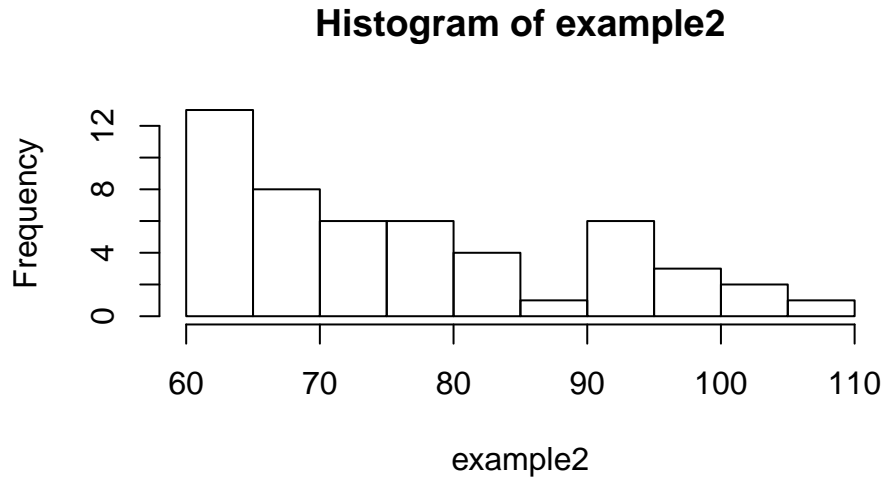
Which measures of location and spread should be used for data that is distributed as shown in this histogram?



SOLUTION: The data looks roughly normally distributed, so you would use the mean and standard deviation.

Question 19

Which measures of location and spread should be used for data that is distributed as shown in this histogram?



SOLUTION: The data looks skewed, so you would use the median and IQR.

Question 20

What information is included in the output for the *summary()* command in R?

- a. Minimum, 1st Quartile, Median, Mean, Mode, 3st Quartile, Maximum
- b. 1st Quartile, Median, Mean, 3st Quartile, Standard Deviation, Variance
- c. Minimum, 1st Quartile, Median, Mean, 3st Quartile, Maximum
- d. Minimum, Median, IQR, Mean, Standard Deviation, Maximum

SOLUTION: c.