

Entry quiz

Entry quiz (**not graded**) to see what you already know (if anything) about data science! This course assumes no prior knowledge - the quiz only for me to find out what you already know, and for assessment purposes (you'll get this quiz again at the end). Don't worry if you cannot answer any of the questions - all of this will be taught in the course!

- Questions may have one or more than one correct answer.
- Partial credit is allowed.
- Questions are not timed.

1

Multiple answer 1 point Question

What is the purpose of data science?

- ☒ Decision support
- ☐ Machine learning
- ☐ Data literacy
- ☒ Data visualization

2

Multiple answer 1 point Question

Which of these are skills that data scientists really need?

- ☒ Programming skills
- ☐ Database management
- ☒ Math and statistics
- ☒ Domain knowledge

3

Multiple answer 1 point Question

Which of these things have to do with data science?

- ☐ C++
- ☒ Kaggle
- ☒ SQL
- ☐ Apache HTTP server

4

Multiple answer 1 point Question

What is R (in data science)?

- ☐ Statistical correlation coefficient
- ☒ Statistical programming language
- ☐ Companion package to SQL
- ☒ Free and Open Source Software

5

Multiple choice 1 point Question

R is among the top 20 most popular programming languages (according to the TIOBE ranking)

- ☒ True
- ☐ False

6

Multiple choice 1 point Question

Which language is used to query relational databases?

- ☐ C#
- ☐ R
- ☐ MySQL
- ☒ SQL

7

Multiple answer 1 point Question

What is the logarithm with base 10 of 1000?

- ☒ 0.03×10^2
- ☐ 2
- ☐ 10^{**3}
- ☒ 3

8

Multiple answer 1 point Question

What is $e^{\ln(2.71)}$?

- ☒ $e \approx 2.71$
- ☐ $\ln(e)$
- ☐ $e^{2.71}$
- ☒ $\ln(e^{2.71})$

9

Multiple choice 1 point Question

Which of these numbers is equal to 0.134×10^{-6} ?

- ☐ 1.34
- ☐ 134,000,000
- ☐ $134/10000$
- ☒ $134 \cdot 10^{-9}$

10

Multiple answer 1 point Question

Which of these equations is for a line with the same slope as $y = -3x + 2$? (More than one answer is correct.)

- ☒ $y = -3x - 8$
- ☐ $y = 8x - 3$
- ☐ $y = 5x - 8$
- ☒ $2y = -6x + 4$

11

Multiple choice 1 point Question

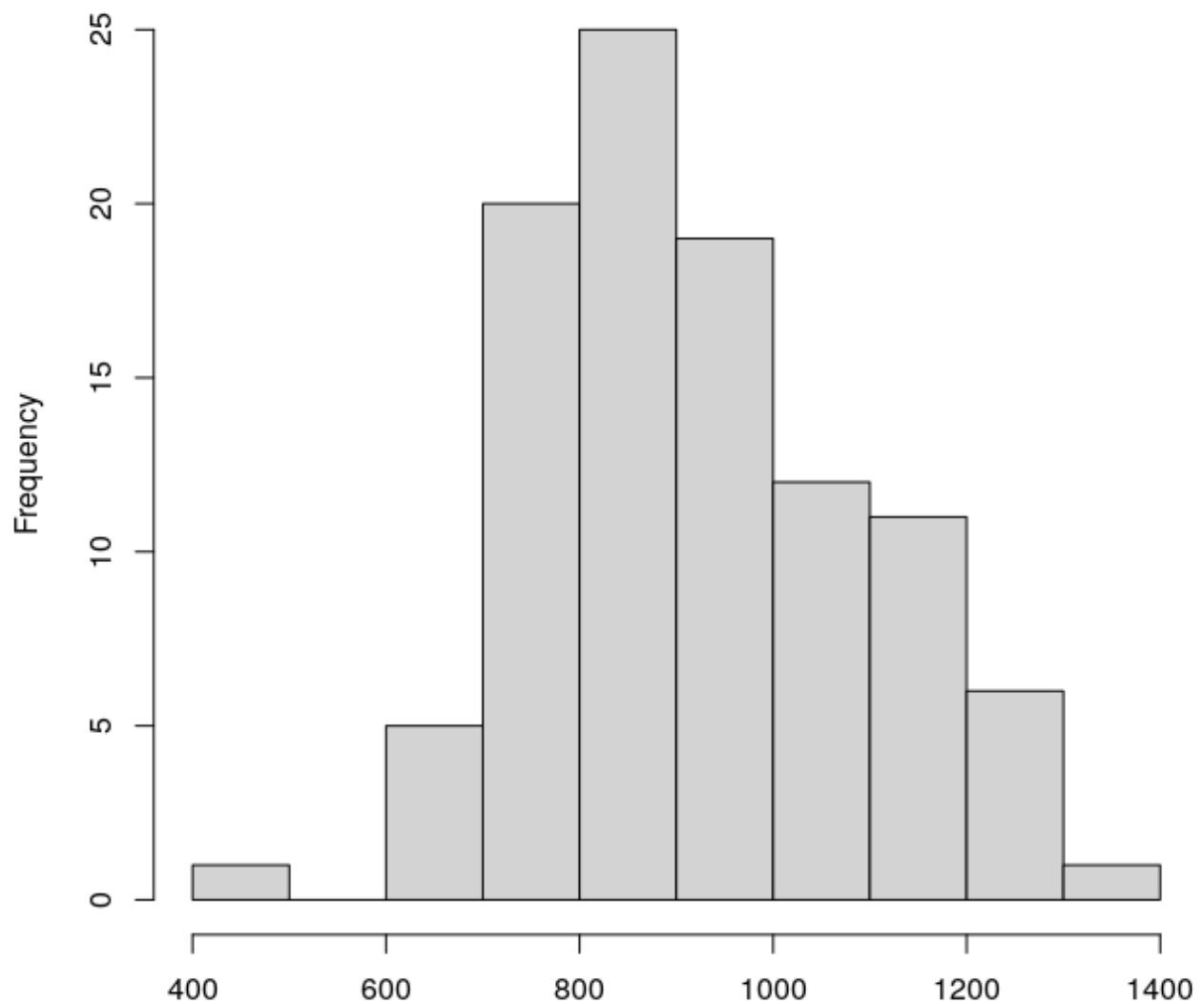
What is the "Git" in "GitHub"?

- ☐ The British word for "big dummy"
- ☐ The German word for "grid" (in German: "Gitter")
- ☒ The version control system created by Linus Torvalds
- ☐ From "GIT" for G[NU] I[s] T[errific]

What does "iterate programming" stand for?

- ☐ Programming for fiction writers
- ☒ A programming paradigm of combining documentation and source code
- ☒ A programming methodology to improve readability of code
- ☐ A structured database querying language style

What type of plot is shown in the picture?



- ☐ Boxplot
- ☐ Heatmap
- ☐ Scatterplot
- ☒ Histogram

14

Multiple choice 1 point Question

This code prints the reciprocal of **a**, unless **a = 0**:

```
if(a!=0){  
    print(1/a)  
} else{  
    print("No reciprocal for 0.")  
}
```

- ☒ True
- ☐ False

15

Multiple choice 1 point Question

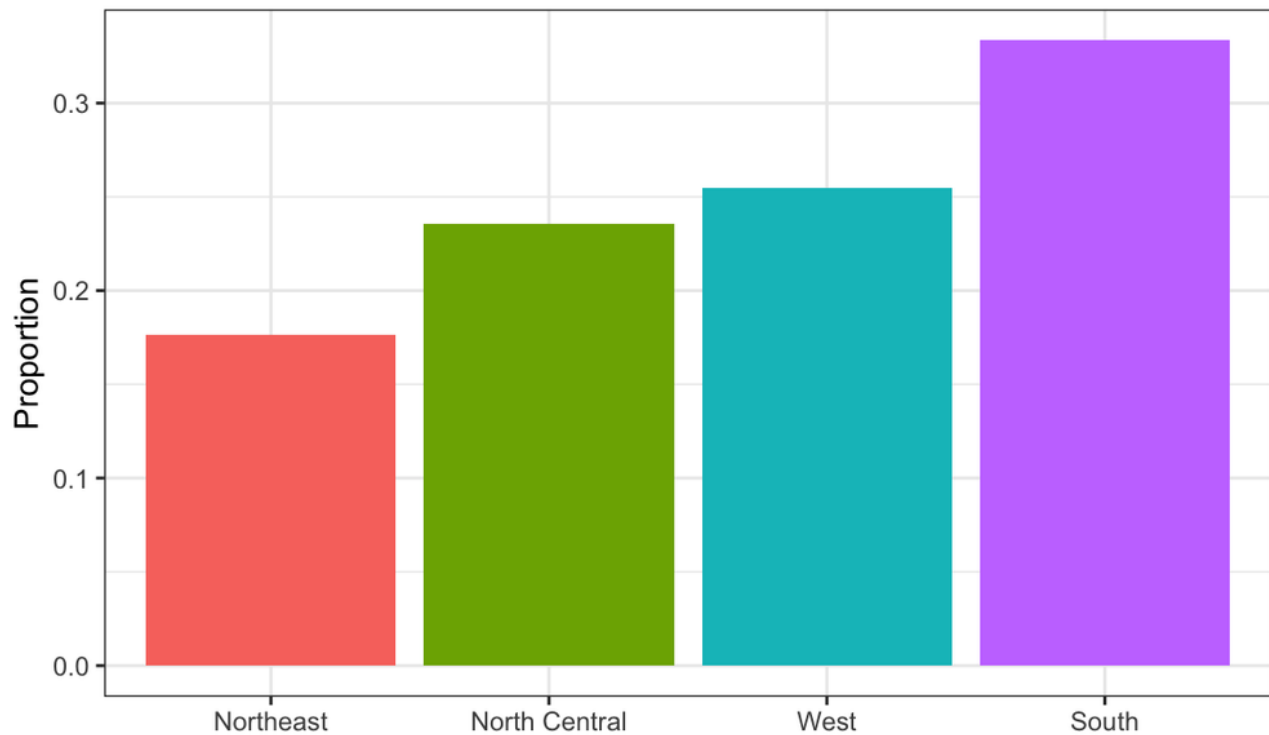
(TRUE, FALSE, TRUE) is a vector of string values.

- ☐ True
- ☒ False

What is a Euclidean vector?

- ☐ An arrow in space
- ☐ A geometric object with an angle in the plane
- ☒ A geometric object with a magnitude and a direction
- ☐ An n-tuple of elements

What kind are the variables on the x axis in the picture?



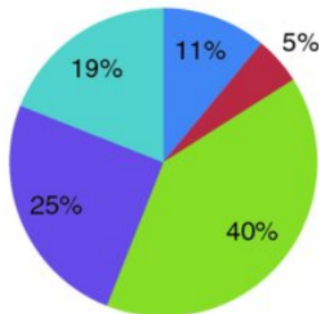
- ☐ Numerical variables
- ☒ Categorical variables
- ☐ Logical variables
- ☐ NaN ("Not a Number")

How efficient is this data presentation?

Survey Results

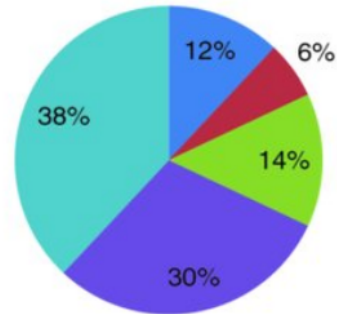
PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



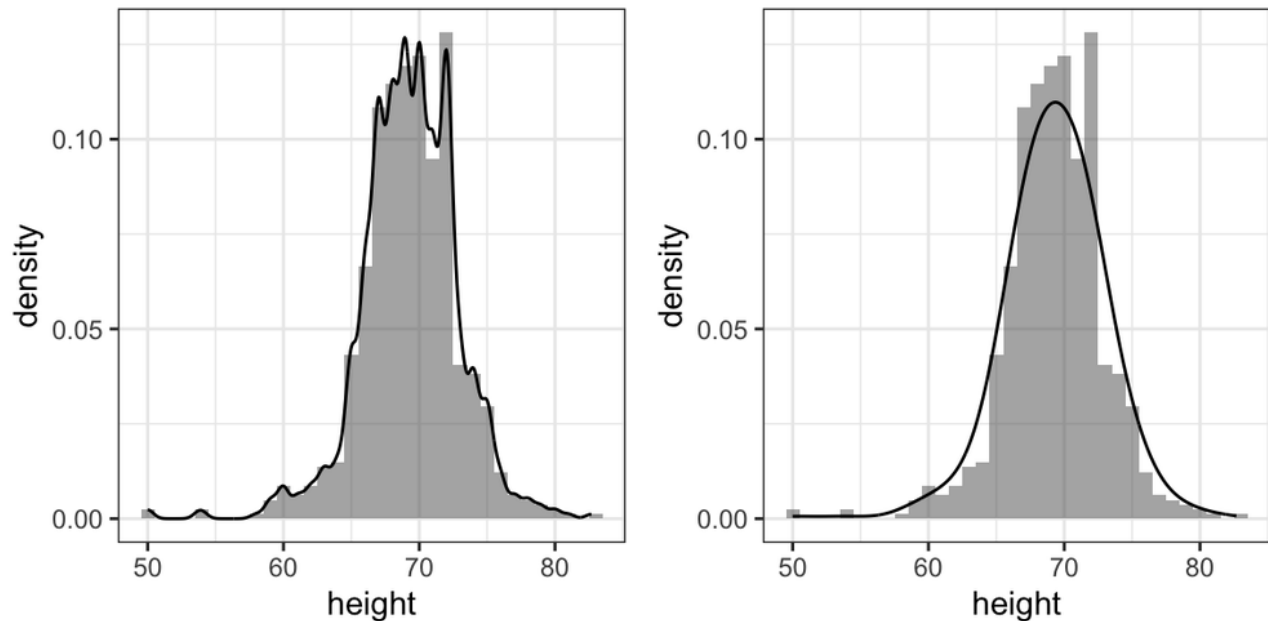
POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



- ☐ Efficient, because of the color scheme
- ☒ Not efficient - hard to compare both charts
- ☐ Depends on the observer
- ☐ Wrong question - efficiency is not relevant

Smoothed density plots (see figure) are based on assumptions about the underlying dataset.



- ☒ True
- ☐ False

Which of these aspects influence effectiveness of a data visualization?

- ☒ Aesthetics (e.g. color)
- ☒ Scaling (e.g. logarithmic)
- ☒ Geometry (e.g. plot type)
- ☒ Data (e.g. choice of variables to be plotted)