Data Carpentry Post-Workshop Assessment

Consent to Participate in Research
Data Carpentry Post-Workshop Assessment

Introduction and Purpose
My name is Erin Becker and I am the Associate Director of Data Carpentry. Thank you for volunteering to take part in our research study, which is about understanding the effectiveness of our workshops. To participate in the study, you will complete a short survey about your skills and attitudes related to our workshop content before and after your workshop. The survey will be 25 questions long and will take approximately 15-20 minutes to complete.

Confidentiality
Your responses will be recorded anonymously. If you respond via email, your IP address will be registered; however, your responses will remain anonymous.

Risks and Benefits
There are no direct risks or benefits to you from filling out this survey, and no compensation. We hope to use these results to improve workshops for future learners.

Consent
You are not required to take this survey to participate in our workshop. You may quit the survey at any time or skip any item other than those required to correctly sort your responses.

If you have any questions about the study, please contact Erin Becker, Associate Director of Data Carpentry at ebecker@datacarpentry.org or eribecker@ucdavis.edu or Megan Welsh, Assistant Professor of Education at the University of California, Davis at megwelsh@ucdavis.edu.

If you have any questions about your rights or treatment as a research participant in this study or would like to provide input about this research, please contact the University of California at Davis' Institutional Review Board (IRB) at (916) 703-9151, IRBAdmin@ucdmc.ucdavis.edu, or 2921 Stockton Blvd, Suite 1400, Room 1429, Sacramento, CA 95817.

* 1. I consent to taking this survey.

- ◯ Yes

* 2. Are you 18 years of age or above?

- ◯ Yes
- ◯ No

3. Which workshop did you attend?

- ◯ Ecology
- ◯ Genomics
- ◯ Geospatial
- ◯ Reproducible Research
- ◯ Social Sciences
- ◯ I don’t know.

* 4. Which workshop did you attend?

* 5. Please enter a unique identifier as follows: Number of siblings (as numeric) + First two letters of the city you were born in (lowercase) + First three letters of your current street (lowercase). This identifier will be confidential to you and will help us pair your results with the post-assessment.

Example: If I have 4 siblings, was born in Arlington, and live on Creekwater Street, my unique identifier would be **4arcre**

* 6. Which of the following programming languages was covered in your workshop?

- R
- Python
- Neither
- I don't know./I don't remember.

Skills Assessment - R

The purpose of this section is to assess your knowledge of the tools you will learn in your workshop. No prior knowledge of these tools is expected of you to participate in this workshop. This is a way for us to understand your knowledge of the tools. In your workshop we will cover all of the skills you see below. If you do not feel comfortable completing this section, please leave these questions blank and continue on to question 14. If the concepts below already make sense to you, you may consider becoming a workshop helper.

7. Which of the following are fundamental rules for producing well formatted spreadsheet tables? Check all that apply.

- [ ] Put each variable (e.g. 'weight' or 'temperature') in its own column.
- [ ] Put each observation in its own row.
- [ ] Combine related pieces of information in one cell.
- [ ] Leave the raw data raw and make edits to a copy of the data.
- [ ] Place comments alongside data values within a single cell, so they don't get separated.

8. The following spreadsheet table shows data from a survey of teenagers' favorite fruit. Multiple researchers have entered data into the spreadsheet keeping track of date collected, school code, age, sex, and favorite fruit.

Date Collected	School Code	Age-Sex	Favorite Fruit
1/19/17	01	15-M	orange
1/19/17	01	17-F	apple
1/19/17	01	18-F	grapes
1/20/17	01	16-F	banana
1/20/17	02	14-M	pear
1/20/17	02	17-F	mango
3/13/17	02	15-F	kiwi
3/13/17	02	18-F	peach
3/13/17	02	16-F	strawberries

Which of the following tables most improves the structure of this data?

Year	Month	Day	School code	Age-Sex	Favorite Fruit
2017	01	09	01	15-M	orange
2017	01	09	01	17-F	apple
2017	01	09	01	18-F	grapes
2017	01	20	01	16-F	banana
2017	01	20	02	14-M	pear
2017	01	20	02	17-F	mango
2017	03	13	02	15-F	kiwi
2017	03	13	02	18-F	peach
2017	03	13	02	16-F	strawberries

Year	Month	Day	School code	Age	Sex	Favorite Fruit
2017	01	09	01	15	M	orange
2017	01	09	01	17	F	apple
2017	01	09	01	18	F	grapes
2017	01	20	01	16	F	banana
2017	01	20	02	14	M	pear
2017	01	20	02	17	F	mango
2017	03	13	02	15	F	kiwi
2017	03	13	02	18	F	peach
2017	03	13	02	16	F	strawberries

Date collected	School code	Age	Sex	Favorite Fruit
1/9/17	01	15	M	orange
1/9/17	01	17	F	apple
1/9/17	01	18	F	grapes
1/20/17	01	16	F	banana
1/20/17	02	14	M	pear
1/20/17	02	17	F	mango
3/13/17	02	15	F	kiwi
3/13/17	02	18	F	peach
3/13/17	02	16	F	strawberries

None of the tables above improve the structure of the data.

9. You collected data in a spreadsheet program and would now like to read your data into R. First, you export the data to a file named "data.txt". This file is shown below.

contact,level,domain,affiliated
"Linda Ramirez<linda.ramirez@gmail.com>",2,”High performance computing”,TRUE
"Trevor Jones <tjones178@ucsf.edu>",1,”Library and information science”,FALSE
"Areej Ahmed <a_ahmed@me.com>",1,"Planetary sciences (geology, climatology)",TRUE

You know that you want to read this data into R using either the read.table or the read.csv functions. The relevant parts of the help files for these two functions are shown below:

read.table(file, header = FALSE, sep = "", ...)
read.csv(file, header = TRUE, sep = "," ...)

How can you read this data into R, creating the dataframe ‘contacts’, so that you can work with the data in R?

- A: contacts <- read.csv(“data.txt”)
- B: contacts <- read.csv(“data.txt”, header = TRUE, sep = “,”)
- C: contacts <- read.table(“data.txt”)
- D: contacts <- read.table(“data.txt”, header = TRUE, sep = “,”)
- Options A and C will both work.
- Options A, B and D will all work.

10. Which of the following options complete the blanks in the statement below to make a true statement? Check all that apply.

Answer A in the previous question _____, because _____.

- [ ] will work, the data is a csv file with headers
- [ ] will NOT work, you need to specify options for all parameters to a function
- [ ] will NOT work, the data isn’t a csv file
- [ ] will NOT work, the data doesn’t have headers

11. After you load data into a dataframe, what are some things you can do to check that it was imported correctly? Check all that apply.

- [ ] Use the str() function to see information about the data.
- [ ] Use the head() function to see the last few lines of the data.
- [ ] Type the name of the data frame to display the whole dataset.
- [ ] Use the dim() function to see the number of rows and columns in the dataset.

12. ggplot is an R package that is used to build plots from data in a dataframe. If ‘df’ is your dataframe and has columns x and y, which of the following lines of code will produce a plot of x versus y?

- ( ) A: ggplot <- df
- ( ) B: ggplot(df)
- ( ) C: ggplot(df, aes(x, y))
- ( ) D: ggplot(df, aes(x, y)) + geom_point()
- ( ) None of the above will work.

13. Which of the following options complete the blanks in the statement below to make a true statement? Check all that apply.

Answer C in the previous question _____, because _____.

- [ ] will work, it contains all of the necessary information.
- [ ] will NOT work, you need to specify the type of plot that you want.
- [ ] will NOT work, you need to specify a color for the points in your plot.
- [ ] will NOT work, you need to specify a size for the points in your plot.

Skills Assessment - Python

The purpose of this section is to assess your knowledge of the tools you will learn in your workshop. No prior knowledge of these tools is expected of you to participate in this workshop. This is a way for us to understand your knowledge of the tools. In your workshop we will cover all of the skills you see below. If you do not feel comfortable completing this section, please leave these questions blank and continue on to question 14. If the concepts below already make sense to you, you may consider becoming a workshop helper.

14. Which of the following are fundamental rules for producing well formatted spreadsheet tables? Check all that apply.

- [ ] Put each variable (e.g. 'weight' or 'temperature') in its own column.
- [ ] Put each observation in its own row.
- [ ] Combine related pieces of information in one cell.
- [ ] Leave the raw data raw and make edits to a copy of the data.
- [ ] Place comments alongside data values within a single cell, so they don't get separated.

15. The following spreadsheet table shows data from a survey of teenagers' favorite fruit. Multiple researchers have entered data into the spreadsheet keeping track of date collected, school code, age, sex, and favorite fruit.

Date Collected	School Code	Age-Sex	Favorite Fruit
1/19/17	01	15-M	orange
1/19/17	01	17-F	apple
1/19/17	01	18-F	grapes
1/20/17	01	16-F	banana
1/20/17	02	14-M	pear
1/20/17	02	17-F	mango
3/13/17	02	15-F	kiwi
3/13/17	02	18-F	peach
3/13/17	02	16-F	strawberries

Which of the following tables most improves the structure of this data?

Year	Month	Day	School code	Age-Sex	Favorite Fruit
2017	01	09	01	15-M	orange
2017	01	09	01	17-F	apple
2017	01	09	01	18-F	grapes
2017	01	20	01	16-F	banana
2017	01	20	02	14-M	pear
2017	01	20	02	17-F	mango
2017	03	13	02	15-F	kiwi
2017	03	13	02	18-F	peach
2017	03	13	02	16-F	strawberries

Year	Month	Day	School code	Age	Sex	Favorite Fruit
2017	01	09	01	15	M	orange
2017	01	09	01	17	F	apple
2017	01	09	01	18	F	grapes
2017	01	20	01	16	F	banana
2017	01	20	02	14	M	pear
2017	01	20	02	17	F	mango
2017	03	13	02	15	F	kiwi
2017	03	13	02	18	F	peach
2017	03	13	02	16	F	strawberries

Date collected	School code	Age	Sex	Favorite Fruit
1/9/17	01	15	M	orange
1/9/17	01	17	F	apple
1/9/17	01	18	F	grapes
1/20/17	01	16	F	banana
1/20/17	02	14	M	pear
1/20/17	02	17	F	mango
3/13/17	02	15	F	kiwi
3/13/17	02	18	F	peach
3/13/17	02	16	F	strawberries

None of the tables above improve the structure of the data.

16. You collected data in a spreadsheet program and would now like to read your data into Python. First, you export the data to a file named "data.txt". This file is shown below.

contact,level,domain,affiliated
"Linda Ramirez<linda.ramirez@gmail.com>",2,”High performance computing”,TRUE
"Trevor Jones <tjones178@ucsf.edu>",1,”Library and information science”,FALSE
"Areej Ahmed <a_ahmed@me.com>",1,"Planetary sciences (geology, climatology)",TRUE

You know that you want to read this data into Python using the pd.read_csv function. The relevant part of the help file for this function is shown below:

Import pandas as pd
pd.read_csv(file)

How can you read this data into Python, creating the dataframe ‘contacts’, so that you can work with the data in Python?

- ( ) A: contacts = pd.read_csv("data.txt")
- ( ) B: contacts = pd.read_csv("data.txt", header = 1, delimiter = ",")
- ( ) C: contacts = pd.read_csv("data.txt", header = 0, delimiter = ",")
- ( ) D: contacts = pd.read_csv("data.txt", header = 1, delimiter = "\t")
- ( ) Options A and C will both work.
- ( ) Options A, B and D will all work.

17. Which of the following options complete the blanks in the statement below to make a true statement? Check all that apply.

Answer A in the previous question _____, because _____.

- [ ] will work, the data is a csv file with headers
- [ ] will NOT work, you need to specify options for all parameters to a function
- [ ] will NOT work, the data isn’t a csv file
- [ ] will NOT work, the data doesn’t have headers

18. After you load data into a dataframe, what are some things you can do to check that it was imported correctly? Check all that apply.

- [ ] Use the type() function to see information about the data
- [ ] Use the head() function to see the last few lines of the data
- [ ] Type the name of the data frame to display the whole dataset
- [ ] Use the contacts.shapes function to see the number of rows and columns in the dataset

19. ggplot is a Python package that is used to build plots from data in a dataframe. If ‘df’ is your dataframe and has columns x and y, which of the following lines of code will produce a plot of x versus y?

- ○ A: ggplot <- df
- ○ B: ggplot(df)
- ○ C: ggplot(df, aes(x, y))
- ○ D: ggplot(df, aes(x, y)) + geom_point()
- ○ None of the above will work.

20. Which of the following options complete the blanks in the statement below to make a true statement? Check all that apply.

Answer C in the previous question _____, because _____.

- ☐ will work, because it contains all of the necessary information.
- ☐ will NOT work, because you need to specify the type of plot that you want.
- ☐ will NOT work, because you need to specify a color for the points in your plot.
- ☐ will NOT work, because you need to specify a size for the points in your plot.

21. Please rate your level of agreement with the following statements:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Having access to the original, raw data is important to be able to repeat an analysis.	○	○	○	○	○
I can write a small program/script/macro to solve a problem in my own work.	○	○	○	○	○
I know how to search for answers to my technical questions online.	○	○	○	○	○
While working on a programming project, if I get stuck, I can find ways of overcoming the problem.	○	○	○	○	○
I am confident in my ability to make use of programming software to work with data.	○	○	○	○	○
Using a programming language (like R or Python) can make my analyses easier to reproduce.	○	○	○	○	○
I can immediately apply what I learned at this workshop.	○	○	○	○	○
I felt comfortable learning in this workshop environment.	○	○	○	○	○
I am confident in my ability to make use of programming languages to work with data.	○	○	○	○	○
Using a programming language (like R or Python) can make me more efficient at working with data.	○	○	○	○	○

22. Please list the major strengths of this workshop?

23. Please list the ways the workshop could be improved?

24. Were there any accessibility issues that affected your ability to participate in this workshop?

- ◯ No
- ◯ Yes

25. If you answered yes to the question above, please describe what the issues were.

Data Carpentry Post-Workshop Assessment

Evaluation of Data Carpentry Instructors

26. Please rate your level of agreement with the following statements:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I was able to get clear answers to my questions from the instructors.	○	○	○	○	○
The instructors were enthusiastic about the workshop.	○	○	○	○	○
I felt comfortable interacting with the instructors.	○	○	○	○	○
The instructors were knowledgeable about the material being taught.	○	○	○	○	○

27. Please provide an example of how an instructor or helper affected your learning experience.

28. How likely are you to recommend this workshop to a friend or colleague?

0 100

* 29. In what country is the workshop you attended being held?

Data Carpentry Post-Workshop Assessment

Demographic Questions

30. What is your gender?

- ◯ Male
- ◯ Female
- ◯ I prefer not to say.

31. How would you describe yourself? Choose one or more of the following groups.

- ☐ American Indian or Alaska Native
  (A person having origins in any of the original peoples of North and South America (including Central America), and who maintains a tribal affiliation or community attachment.)
- ☐ Asian
  (A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.)
- ☐ Black or African American
  (A person having origins in any of the Black racial groups of Africa – includes Caribbean Islanders and others of African origin.)
- ☐ Hispanic or Latino(a)
  (A person of Spanish-speaking origin or ancestry and/or Latin American origin or ancestry – includes Portuguese and Brazilians.)
- ☐ Native Hawaiian or Other Pacific Islander
  (A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.)
- ☐ White
  (A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.)
- ☐ I prefer not to say.
- ☐ Other (please specify)

Thank you for completing this survey. Be sure to check out our blog on www.datacarpentry.com, and follow @datacarpentry on Twitter.