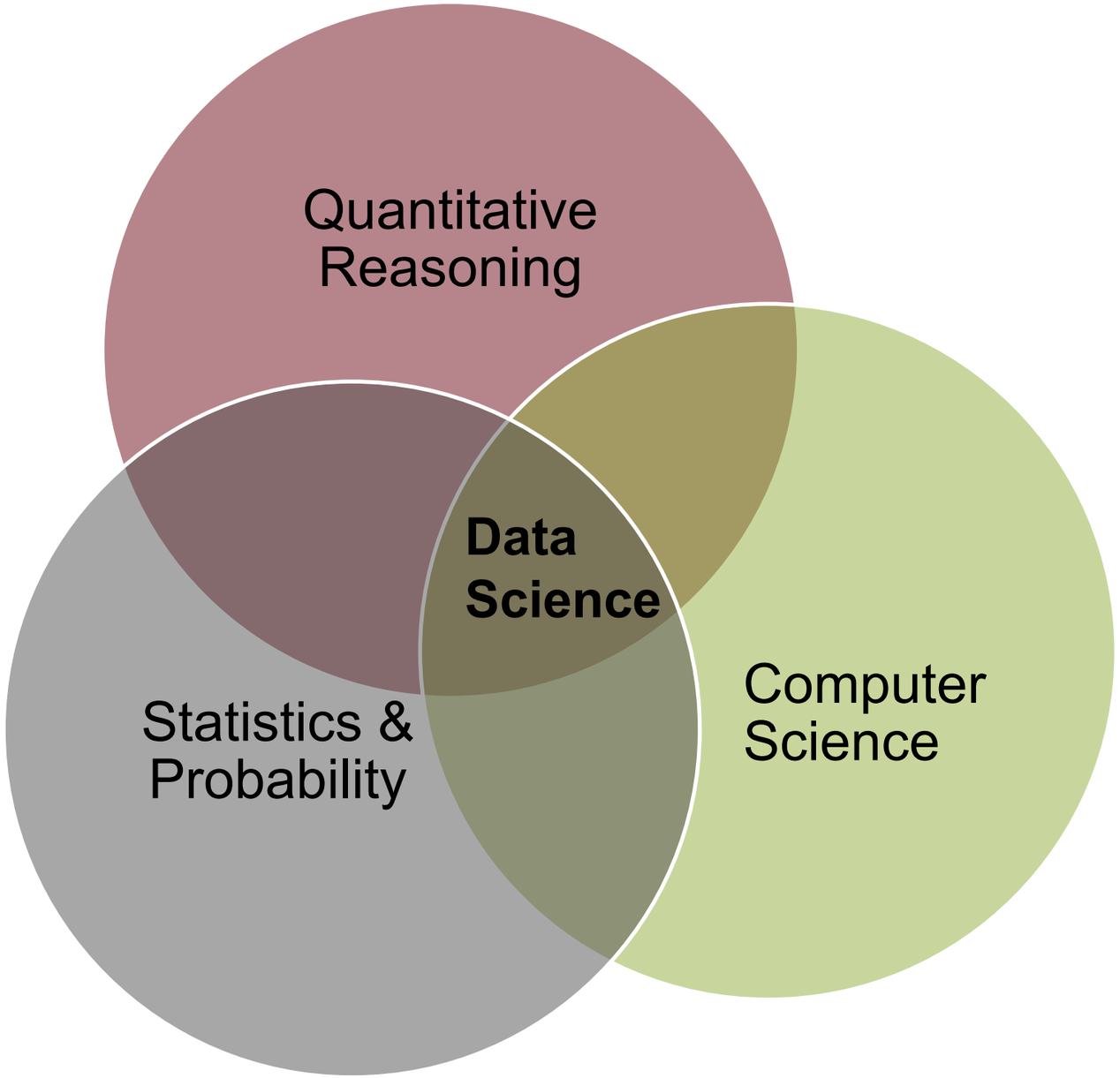
High School Math Pathways Symposium

Data Science Foundations

Nov. 9-10, 2021

What is Data Science?

Data scientists find patterns in sets of information that provide insights about the past, the present and the future.



Data Science

Data science is the intersection of the fields of quantitative reasoning, statistics, and computer science with a heavy focus on the creative problem-solving aspect.

Why is Data Science important?

Postsecondary Pathways

Bachelor's or above (requires Calculus)	Bachelor's or above (requires non-Calculus)	Associate's degree	Certificate/Bootcamps
• Business Intelligence Analysts • Data Science • Information Science • Marketing • Software Developers	• Any Arts or Humanities degree that requires Quantitative Reasoning • Applied Business (AAB) • Applied Marketing (AAB) • Computer Programmers • Criminal Justice (Applied degree) • Database Administrators • Journalism • Market Research Analysts and Marketing Specialists • Management Analysts • Project Managers • Public Relations/Advertising • Others	• Any Arts or Humanities degree that requires Quantitative Reasoning • Computer Network Support Specialists • Web Developers	• Data Analytics • Data Science

Students who are interested in Calc-based careers also need to take Algebra 2.

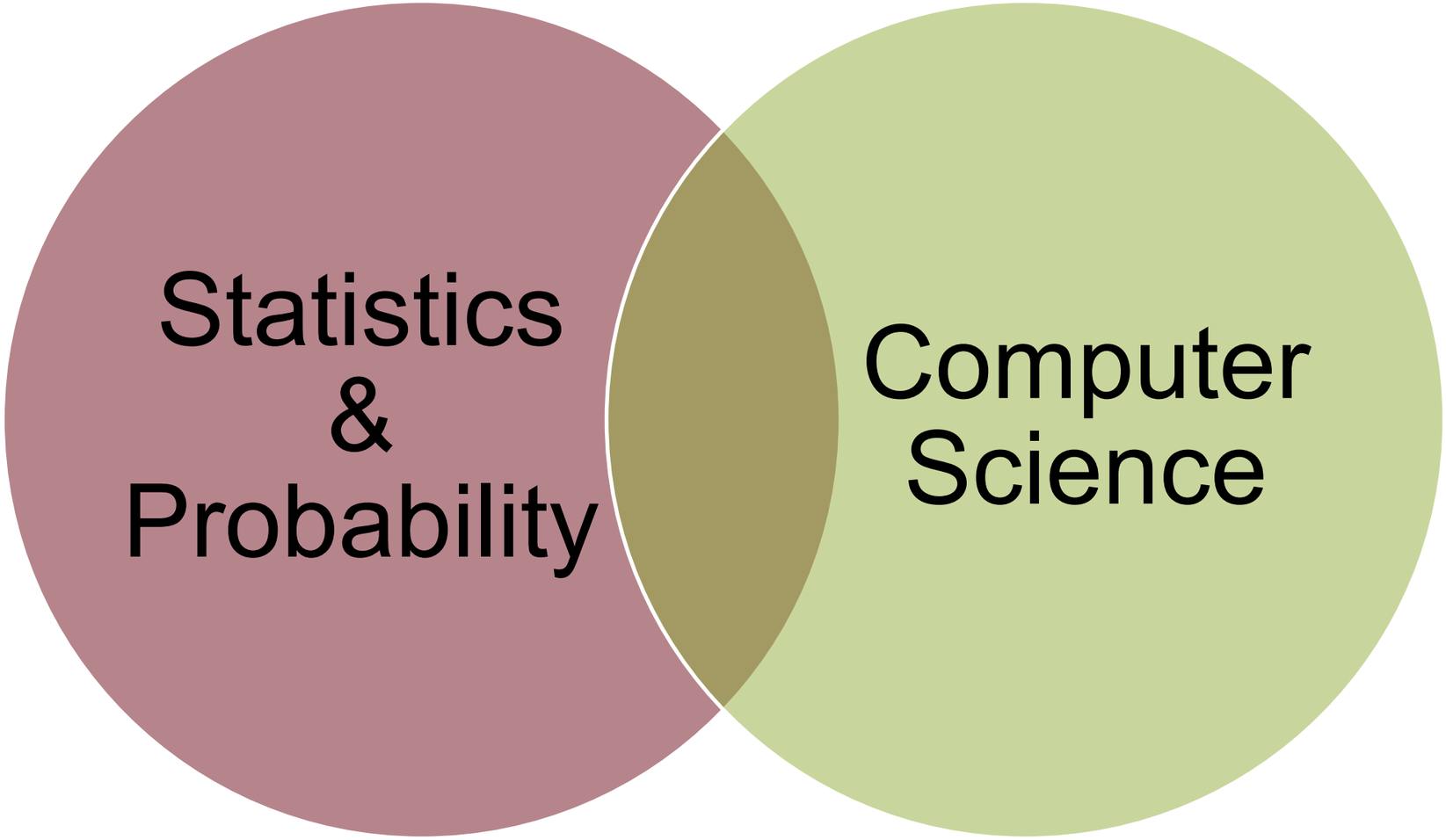
Target Students

Data Science Foundations is beneficial for students who need a third or fourth credit in mathematics and are not intending to pursue a career that requires calculus. It is appropriate for students with limited or no prior programming, statistics and data analytics knowledge. This course is ideal for absolute beginners who want to acquire a basic working knowledge of data science. Data Science Foundations is designed to be a hands-on course that promotes reasoning using the standards for mathematical practice.

The course is especially appropriate for a student who has the following characteristics:

- Anticipates a career in behavioral sciences;
- Anticipates a career in the emerging fields of computer science, computational data analysis and/or statistics;
- Is interested in applied fields of study that use mathematics;
- Enjoys exploring engaging, real-world issues involving data;
- Desires to become a better-informed citizen;
- Plans on pursuing a pathway that does not require calculus; and/or
- Plans on pursuing computer technology or STEM fields at a postsecondary institution.

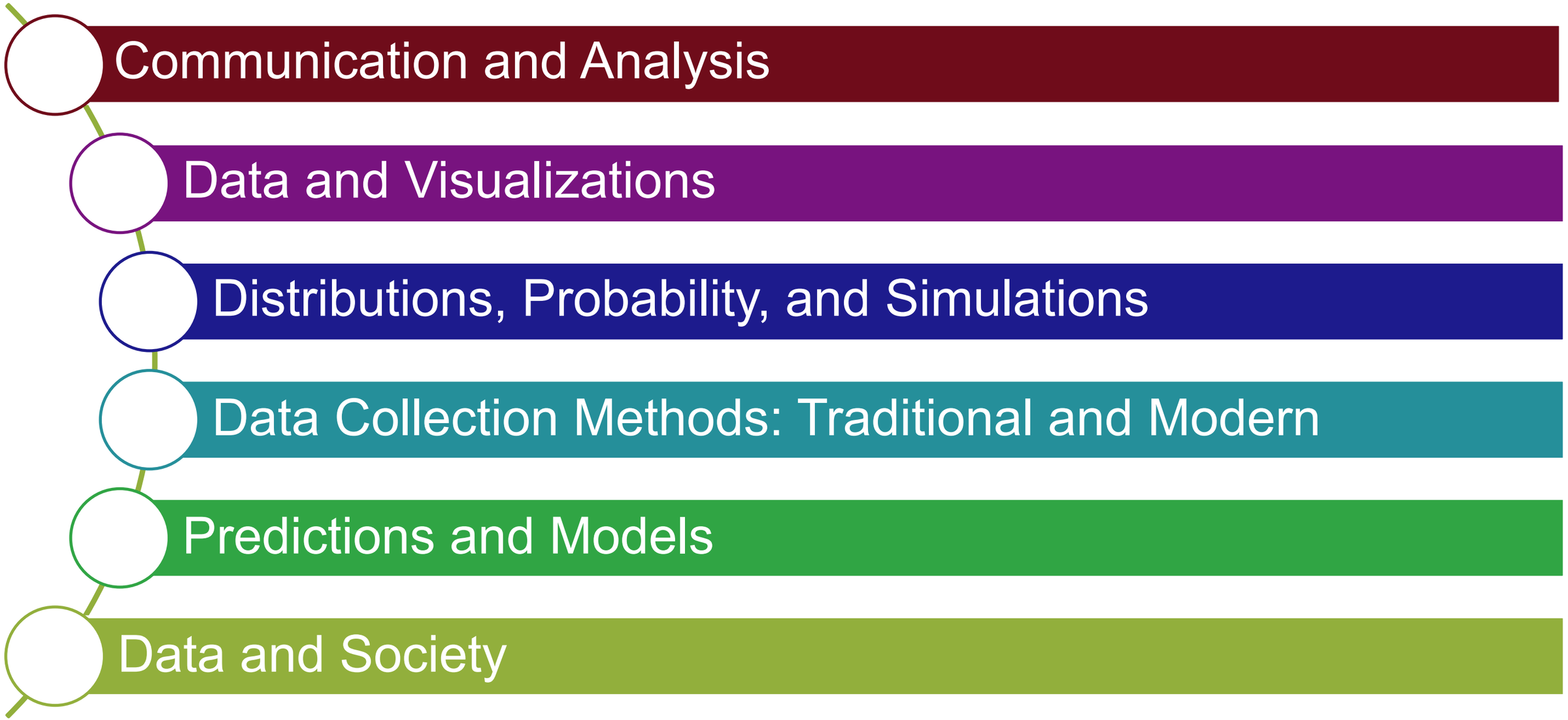
Data Science Standards



Standards Document

https://education.ohio.gov/Topics/Learning-in-Ohio/Mathematics/Resources-for-Mathematics/Math-Pathways/Higher-Ed-Entry-Level-Math-Pathways-Course-Descrip

Overview of Data Science Foundations Course



Overview of Data Science Foundations Course

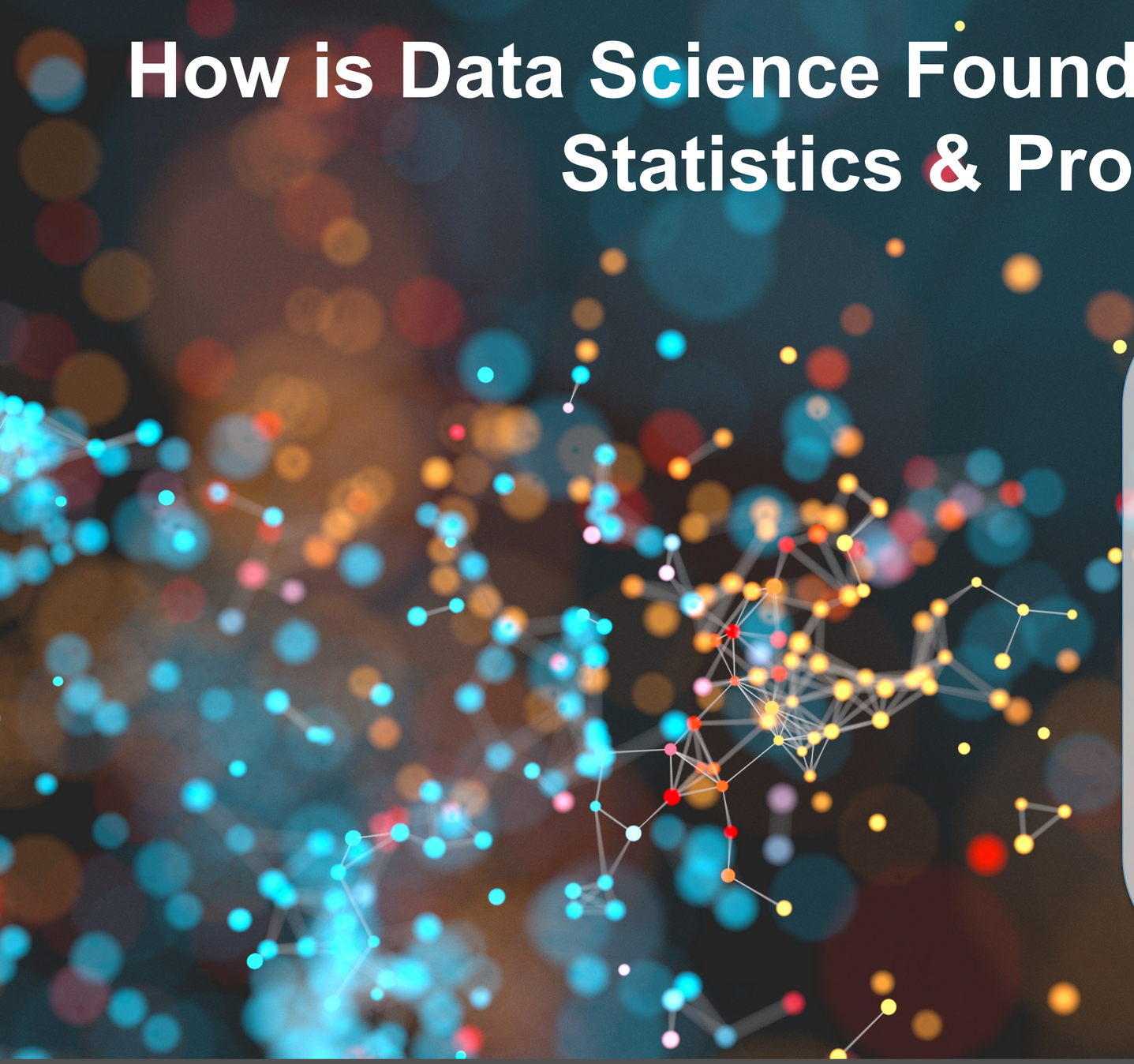
https://www.introdatascience.org/

Overview of Data Science Foundations Course

Technology will be hosted by The University of Cincinnati.

Aligns with emerging postsecondary Introduction to Data Science course

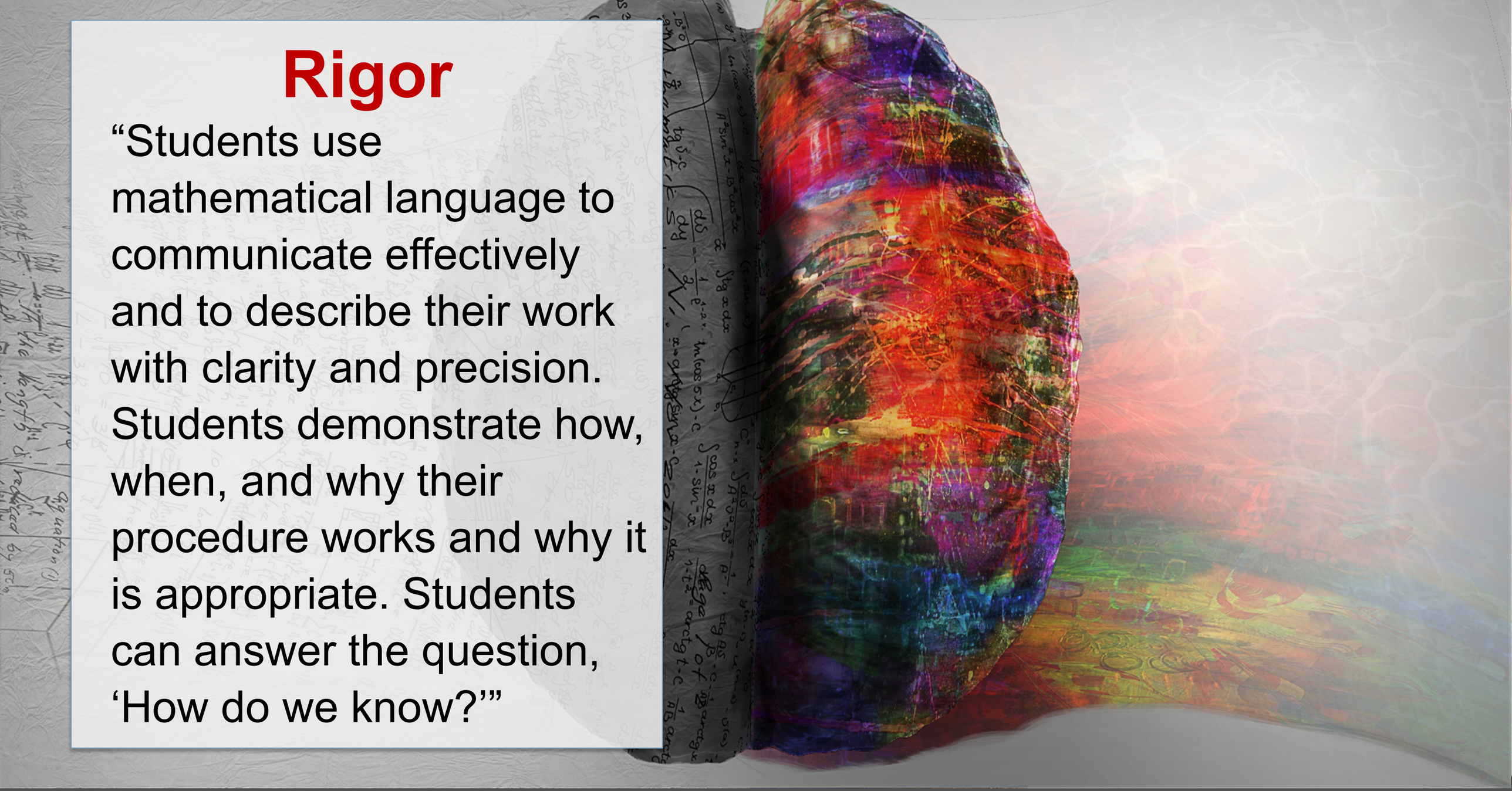
How is Data Science Foundations different than Statistics & Probability?



The big difference between data science and statistics is that where statistics focuses on explaining the data, data science focuses on uncovering insights that help make predictions and decisions.

Rigor

"Students use mathematical language to communicate effectively and to describe their work with clarity and precision. Students demonstrate how, when, and why their procedure works and why it is appropriate. Students can answer the question, 'How do we know?'"



Rigorous courses are…	Rigorous courses are not…
Defined by complexity, which is a measure of the thinking, action, or knowledge that is needed to complete the task	Characterized by difficulty, which is a measure of effort required to complete a task
Measured in depth of understanding	Measured by the amount of work
Opportunities for precision in reasoning, language, definitions, and notation that are sufficient to appropriate age/course	Based on procedure alone
Determined by students' process	Measured by assigning difficult problems
Opportunities for students to make decisions in problem solving	Defined only by the resources used

Rigorous courses are…	Rigorous courses are not…
Opportunities to make connections	Taught in isolation
Supportive of the transfer of knowledge to new situations	Repetitive
Driven by students developing efficient explanations of solutions and why they work, providing opportunities for thinking and reasoning about contextual problems and situations	Focused on getting an answer
Defined by what the student does with what you give them	Defined by what you give the student

Follow-on Courses

- CCP Introduction to Data Science
- AP Statistics
- CCP Introductory Statistics
- CCP Quantitative Reasoning
- CCP Mathematics for Elementary Education
- AP Computer Science A
- Algebra 2
- Other Algebra 2 Equivalent Math Pathways course
- Other CCP math course

Data Science Foundations Pilot

Curriculum

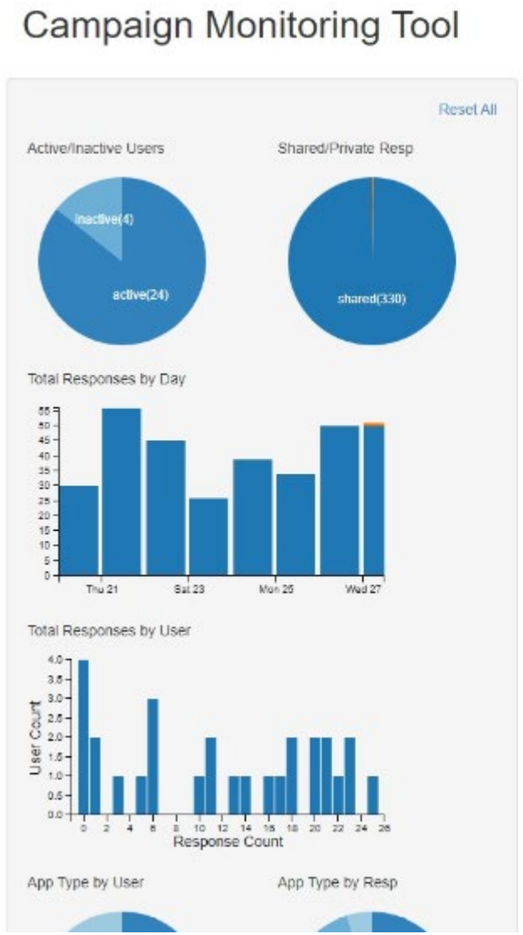
Collaborative Environment

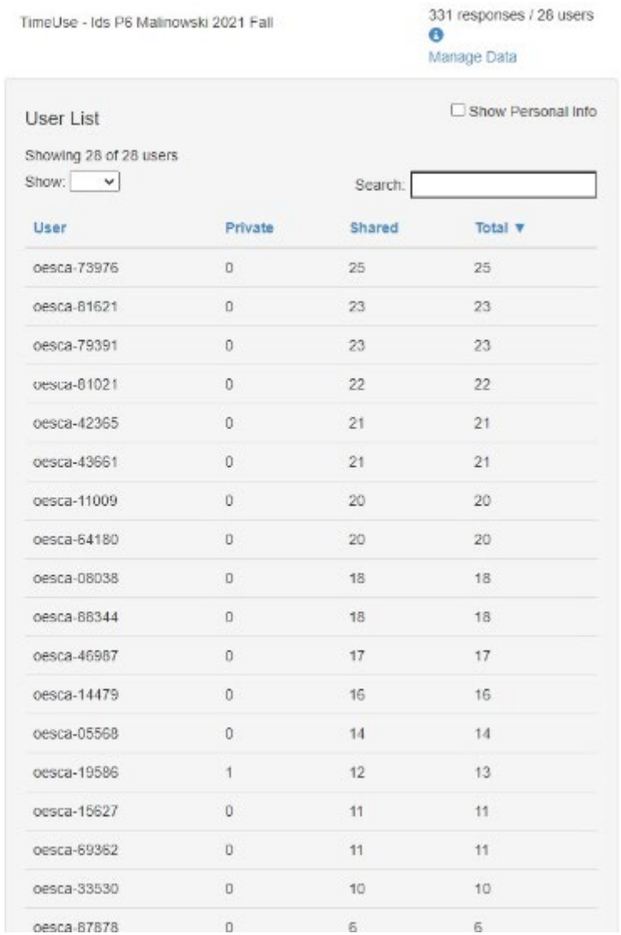
Differentiation

Student-Centered Learning

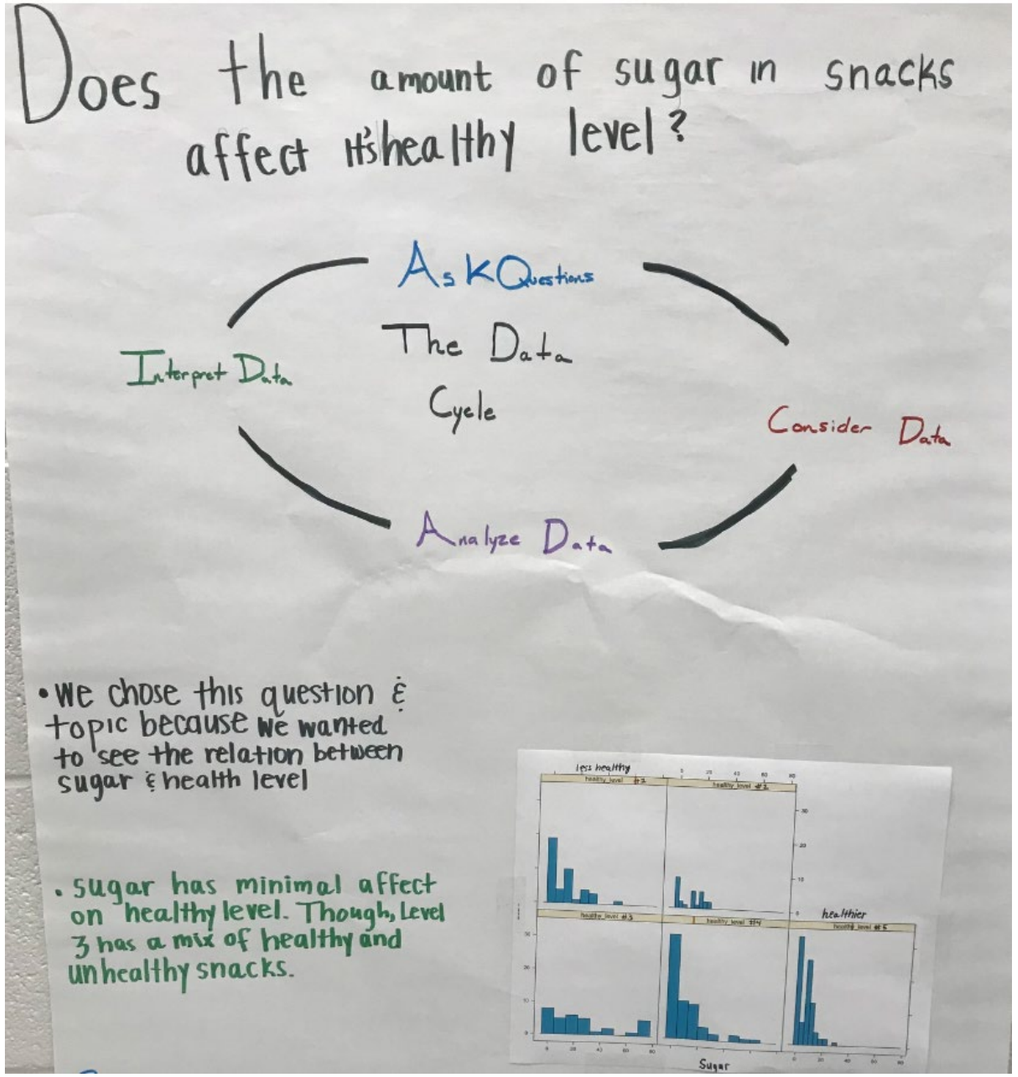
Data Science Foundations Pilot

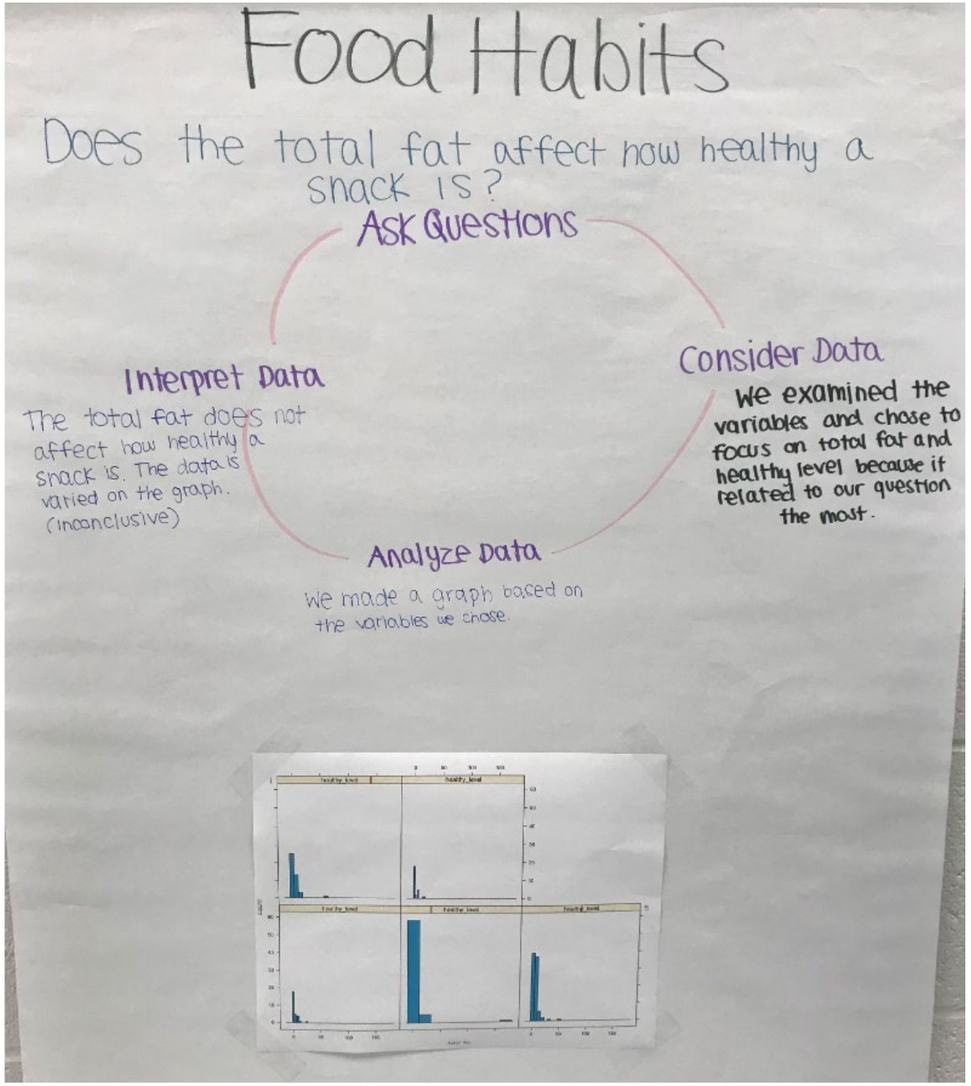
Campaigns
Food Habits
Time Use
Personality Color
Stress/Chill
Class Campaign
Water Campaign



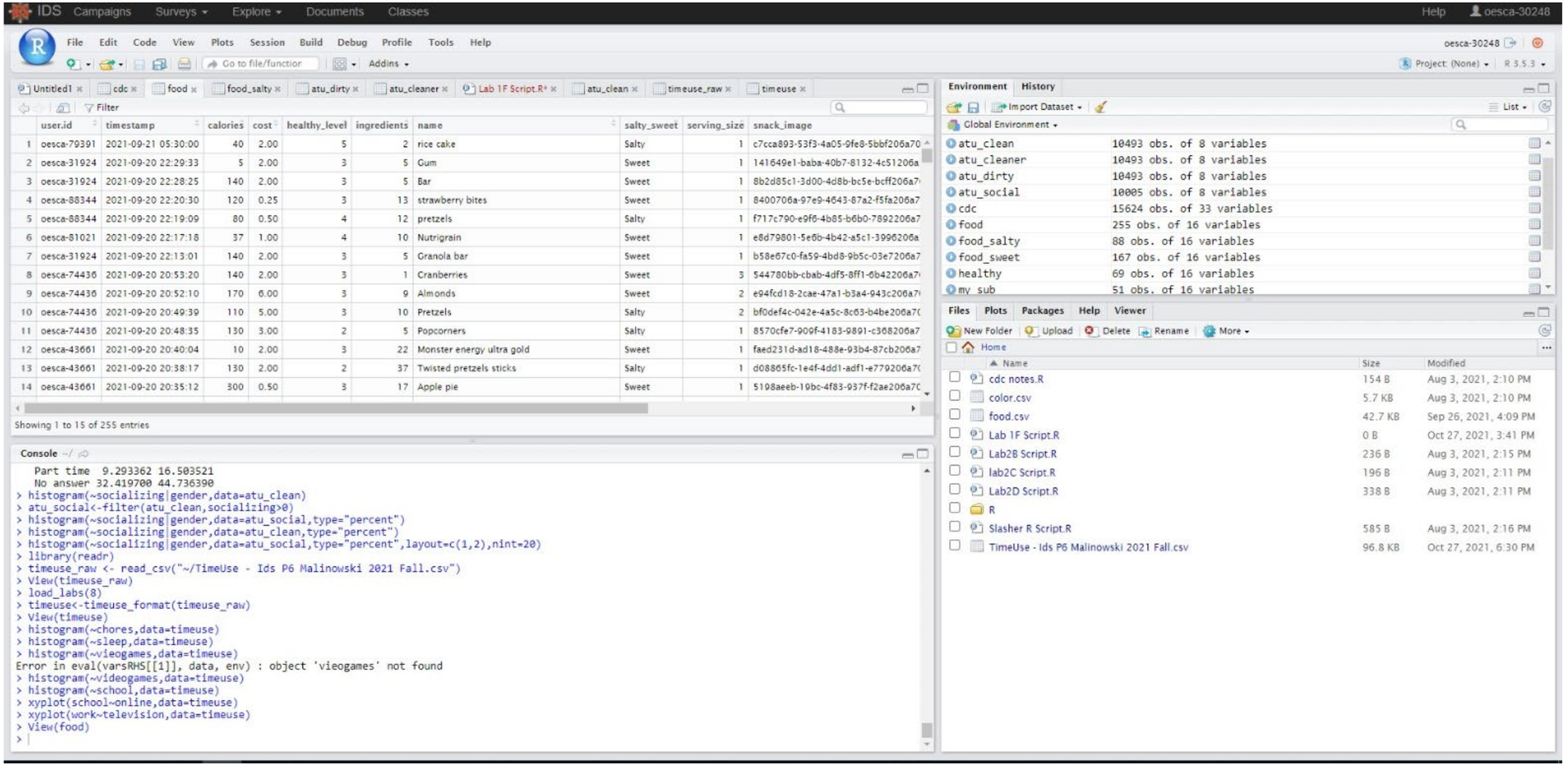


Data Collection Cycle

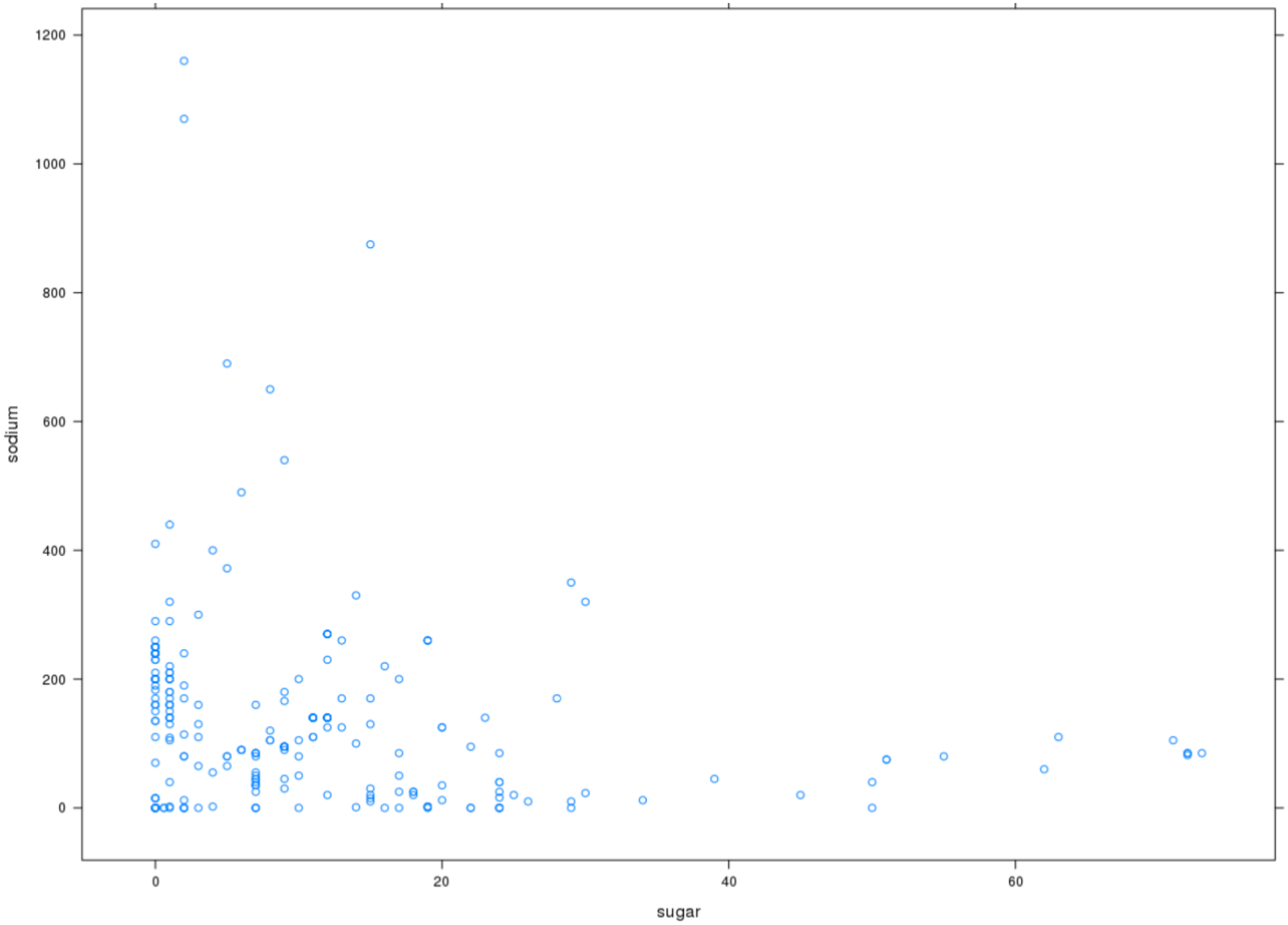




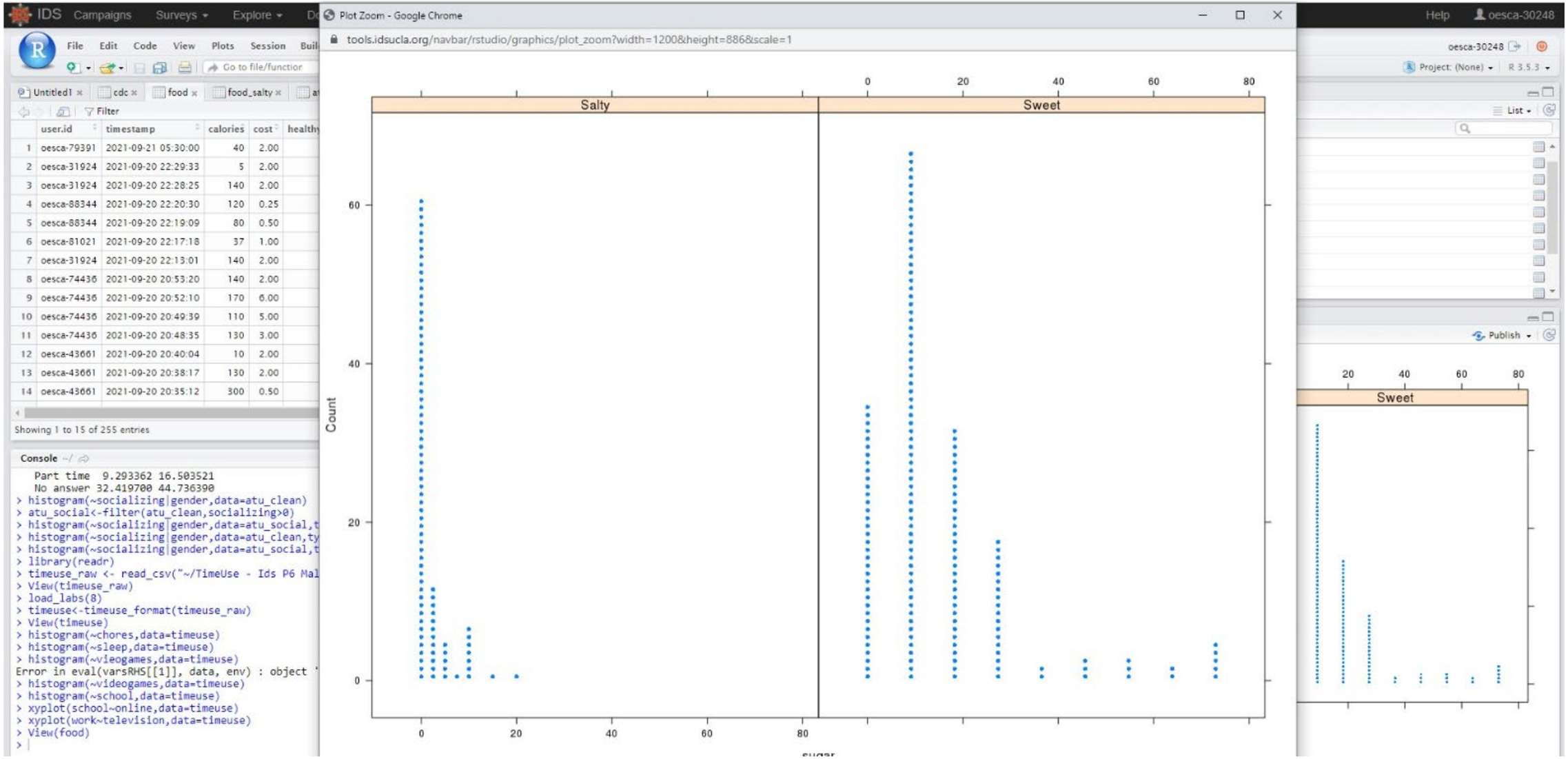
R/RStudio

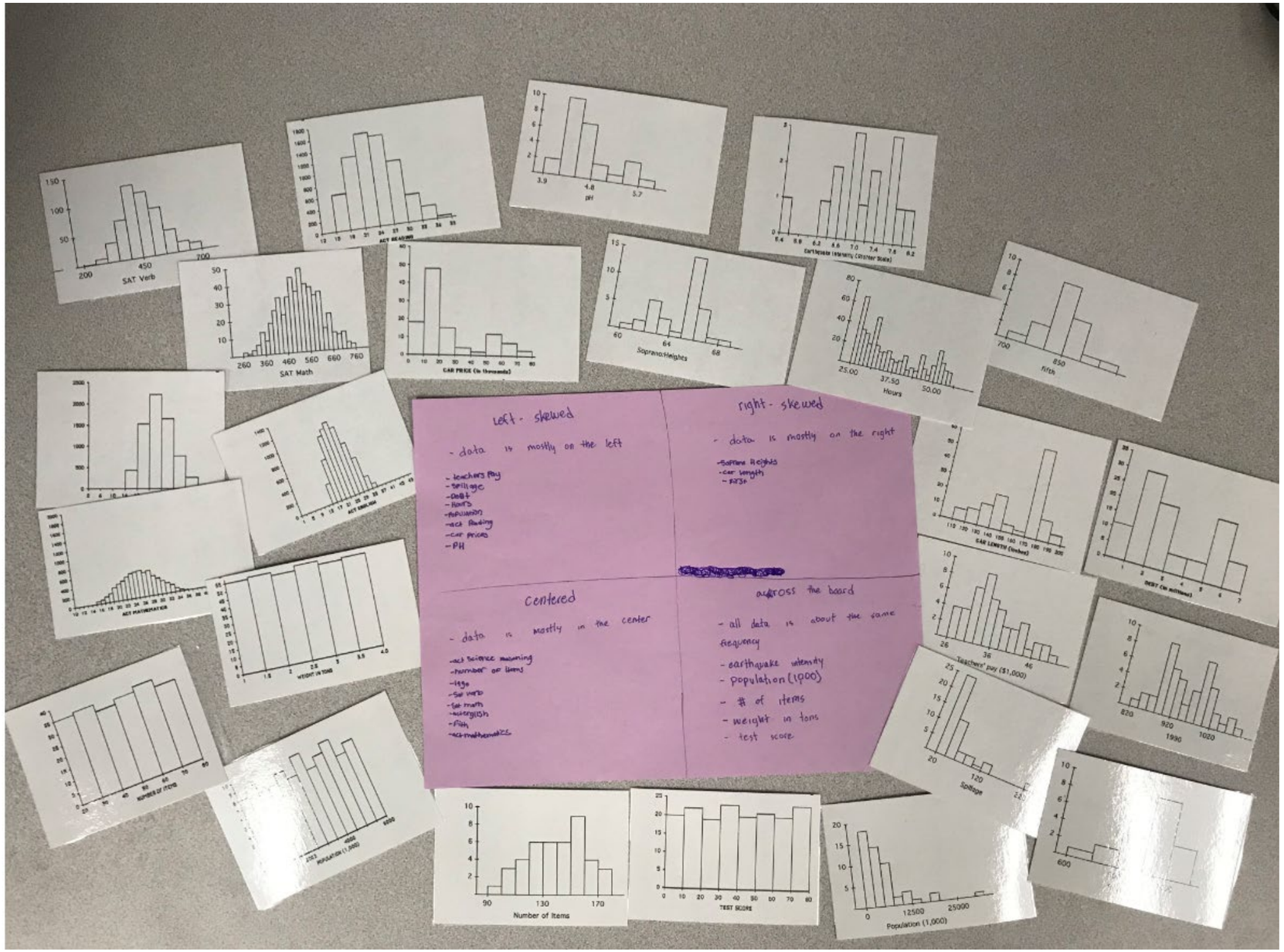

IDS Campaigns Surveys Explore Documents Classes Help oesca-30248
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Project: (None) R 3.5.3
Untitled1 cdc food food_salty atu_dirty atu_cleaner Lab 1F Script.R* atu_clean timeuse_raw timeuse
Filter
user.id timestamp calories cost healthy_level ingredients name salty_sweet serving_size snack_image
1 oesca-79391 2021-09-21 05:30:00 40 2.00 5 2 rice cake Salty 1 c7cca893-53f3-4a05-9fe8-5bbf206a70
2 oesca-31924 2021-09-20 22:29:33 5 2.00 3 5 Gum Sweet 1 141649e1-baba-40b7-8132-4c51206a
3 oesca-31924 2021-09-20 22:28:25 140 2.00 3 5 Bar Sweet 1 8b2d85c1-3d00-4d8b-bc5e-bcff206a7
4 oesca-88344 2021-09-20 22:20:30 120 0.25 3 13 strawberry bites Sweet 1 8400706a-97e9-4643-87a2-f5fa206a7
5 oesca-88344 2021-09-20 22:19:09 80 0.50 4 12 pretzels Salty 1 f717c790-e9f6-4b85-b6b0-7892206a7
6 oesca-81021 2021-09-20 22:17:18 37 1.00 4 10 Nutrigrain Sweet 1 e8d79801-5e6b-4b42-a5c1-3996206a
7 oesca-31924 2021-09-20 22:13:01 140 2.00 3 5 Granola bar Sweet 1 b58e67c0-fa59-4bd8-9b5c-03e7206a7
8 oesca-74436 2021-09-20 20:53:20 140 2.00 3 1 Cranberries Sweet 3 544780bb-cbab-4df5-8ff1-6b42206a7
9 oesca-74436 2021-09-20 20:52:10 170 6.00 3 9 Almonds Sweet 2 e94fcd18-2cae-47a1-b3a4-943c206a7
10 oesca-74436 2021-09-20 20:49:39 110 5.00 3 10 Pretzels Salty 2 bf0def4c-042e-4a5c-8c63-b4be206a7
11 oesca-74436 2021-09-20 20:48:35 130 3.00 2 5 Popcorners Salty 1 8570cfe7-909f-4183-9891-c368206a7
12 oesca-43661 2021-09-20 20:40:04 10 2.00 3 22 Monster energy ultra gold Sweet 1 faed231d-ad18-488e-93b4-87cb206a7
13 oesca-43661 2021-09-20 20:38:17 130 2.00 2 37 Twisted pretzels sticks Salty 1 d08865fc-1e4f-4dd1-adf1-e779206a7
14 oesca-43661 2021-09-20 20:35:12 300 0.50 3 17 Apple pie Sweet 1 5108aeeb-19bc-4f83-937f-f2ae206a7
Showing 1 to 15 of 255 entries
Console ~/
Part time 9.293362 16.503521
No answer 32.419700 44.736390
> histogram(~socializing|gender,data=atu_clean)
> atu_social<-filter(atu_clean,socializing>0)
> histogram(~socializing|gender,data=atu_social,type="percent")
> histogram(~socializing|gender,data=atu_clean,type="percent")
> histogram(~socializing|gender,data=atu_social,type="percent",layout=c(1,2),nint=20)
> library(readr)
> timeuse_raw <- read_csv("~/TimeUse - Ids P6 Malinowski 2021 Fall.csv")
> View(timeuse_raw)
> load_labs(8)
> timeuse<-timeuse_format(timeuse_raw)
> View(timeuse)
> histogram(~chores,data=timeuse)
> histogram(~sleep,data=timeuse)
> histogram(~vieogames,data=timeuse)
Error in eval(varsRHS[[1]], data, env) : object 'vieogames' not found
> histogram(~videogames,data=timeuse)
> histogram(~school,data=timeuse)
> xyplot(school~online,data=timeuse)
> xyplot(work~television,data=timeuse)
> View(food)
Environment History
Import Dataset List
Global Environment
atu_clean 10493 obs. of 8 variables
atu_cleaner 10493 obs. of 8 variables
atu_dirty 10493 obs. of 8 variables
atu_social 10005 obs. of 8 variables
cdc 15624 obs. of 33 variables
food 255 obs. of 16 variables
food_salty 88 obs. of 16 variables
food_sweet 167 obs. of 16 variables
healthy 69 obs. of 16 variables
my_sub 51 obs. of 16 variables
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Home
Name Size Modified
cdc notes.R 154 B Aug 3, 2021, 2:10 PM
color.csv 5.7 KB Aug 3, 2021, 2:10 PM
food.csv 42.7 KB Sep 26, 2021, 4:09 PM
Lab 1F Script.R 0 B Oct 27, 2021, 3:41 PM
Lab2B Script.R 236 B Aug 3, 2021, 2:15 PM
lab2C Script.R 196 B Aug 3, 2021, 2:11 PM
Lab2D Script.R 338 B Aug 3, 2021, 2:11 PM
R
Slasher R Script.R 585 B Aug 3, 2021, 2:16 PM
TimeUse - Ids P6 Malinowski 2021 Fall.csv 96.8 KB Oct 27, 2021, 6:30 PM

R/RStudio



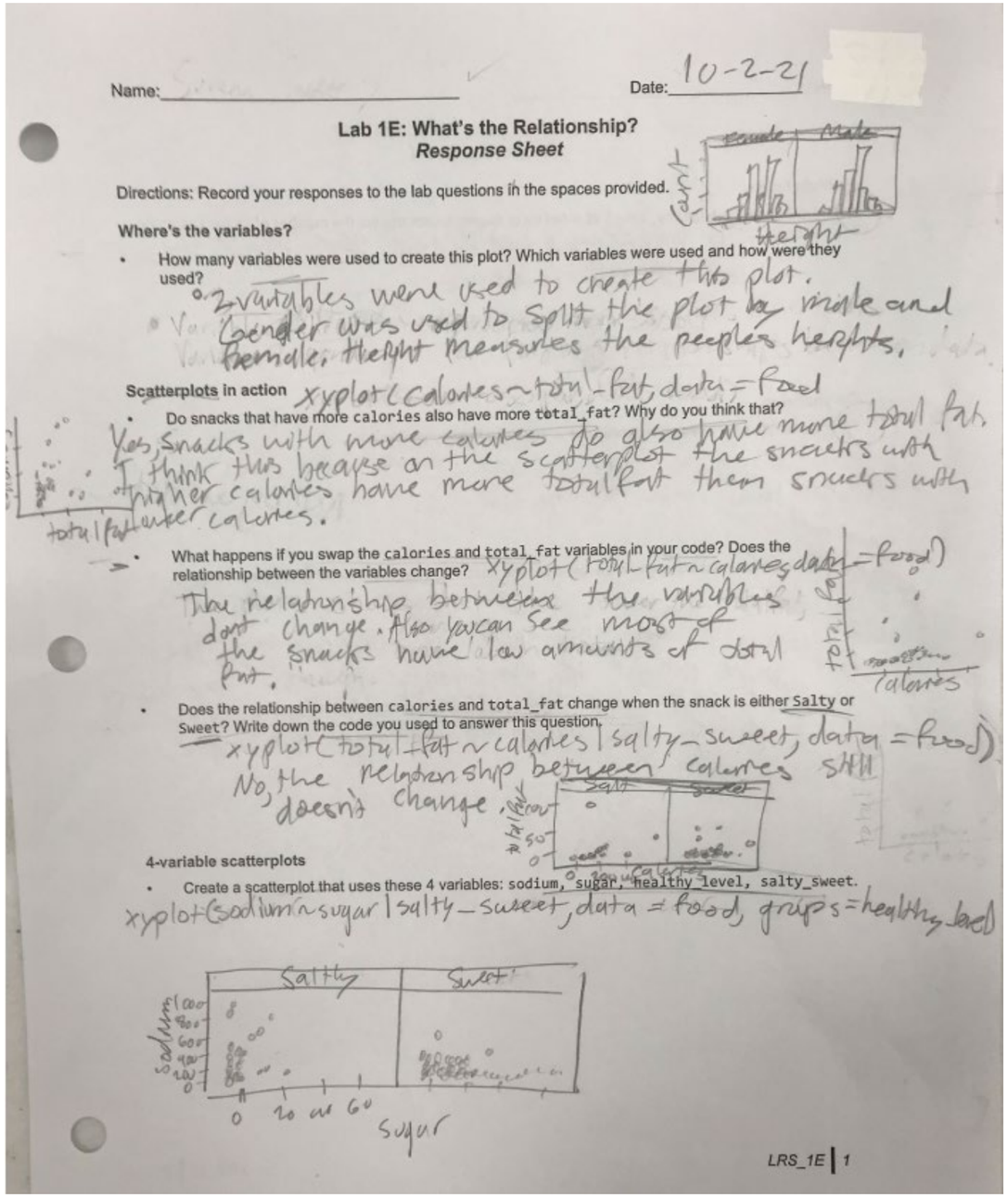
R/RStudio


IDS Campaigns Surveys Explore
File Edit Code View Plots Session
Untitled1 cdc food food_salty
Filter
user.id timestamp calories cost
1 oesca-79391 2021-09-21 05:30:00 40 2.00
2 oesca-31924 2021-09-20 22:29:33 5 2.00
3 oesca-31924 2021-09-20 22:28:25 140 2.00
4 oesca-88344 2021-09-20 22:20:30 120 0.25
5 oesca-88344 2021-09-20 22:19:09 80 0.50
6 oesca-81021 2021-09-20 22:17:18 37 1.00
7 oesca-31924 2021-09-20 22:13:01 140 2.00
8 oesca-74436 2021-09-20 20:53:20 140 2.00
9 oesca-74436 2021-09-20 20:52:10 170 6.00
10 oesca-74436 2021-09-20 20:49:39 110 5.00
11 oesca-74436 2021-09-20 20:48:35 130 3.00
12 oesca-43661 2021-09-20 20:40:04 10 2.00
13 oesca-43661 2021-09-20 20:38:17 130 2.00
14 oesca-43661 2021-09-20 20:35:12 300 0.50
Showing 1 to 15 of 255 entries
Console
Part time 9.293362 16.503521
No answer 32.419700 44.736390
> histogram(~socializing|gender,data=atu_clean)
> atu_social<-filter(atu_clean,socializing>0)
> histogram(~socializing|gender,data=atu_social,t
> histogram(~socializing|gender,data=atu_clean,ty
> histogram(~socializing|gender,data=atu_social,t
> library(readr)
> timeuse_raw <- read_csv("~/TimeUse - Ids P6 Mal
> View(timeuse_raw)
> load_labs(8)
> timeuse<-timeuse_format(timeuse_raw)
> View(timeuse)
> histogram(~chores,data=timeuse)
> histogram(~sleep,data=timeuse)
> histogram(~videogames,data=timeuse)
Error in eval(varsRHS[[1]], data, env) : object '
> histogram(~videogames,data=timeuse)
> histogram(~school,data=timeuse)
> xyplot(school~online,data=timeuse)
> xyplot(work~television,data=timeuse)
> View(food)
Plot Zoom - Google Chrome
tools.idsucla.org/navbar/rstudio/graphics/plot_zoom?width=1200&height=886&scale=1
Salty
Sweet
Count
sugar
Help oesca-30248
Project: (None) R 3.5.3
List
Publish
Sweet



Student Work

Student Work



Name:_____ Date: 10-2-21

Lab 1E: What's the Relationship?
Response Sheet

Directions: Record your responses to the lab questions in the spaces provided.

Where's the variables?

- How many variables were used to create this plot? Which variables were used and how were they used?
  - 2 variables were used to create this plot. Gender was used to split the plot by male and female. Height measures the peeples heights.

Scatterplots in action xyplot(calories~total_fat, data=food)

- Do snacks that have more calories also have more total_fat? Why do you think that?
  Yes, snacks with more calories do also have more total fat. I think this because on the scatterplot the snacks with higher calories have more total fat than snacks with lower calories.

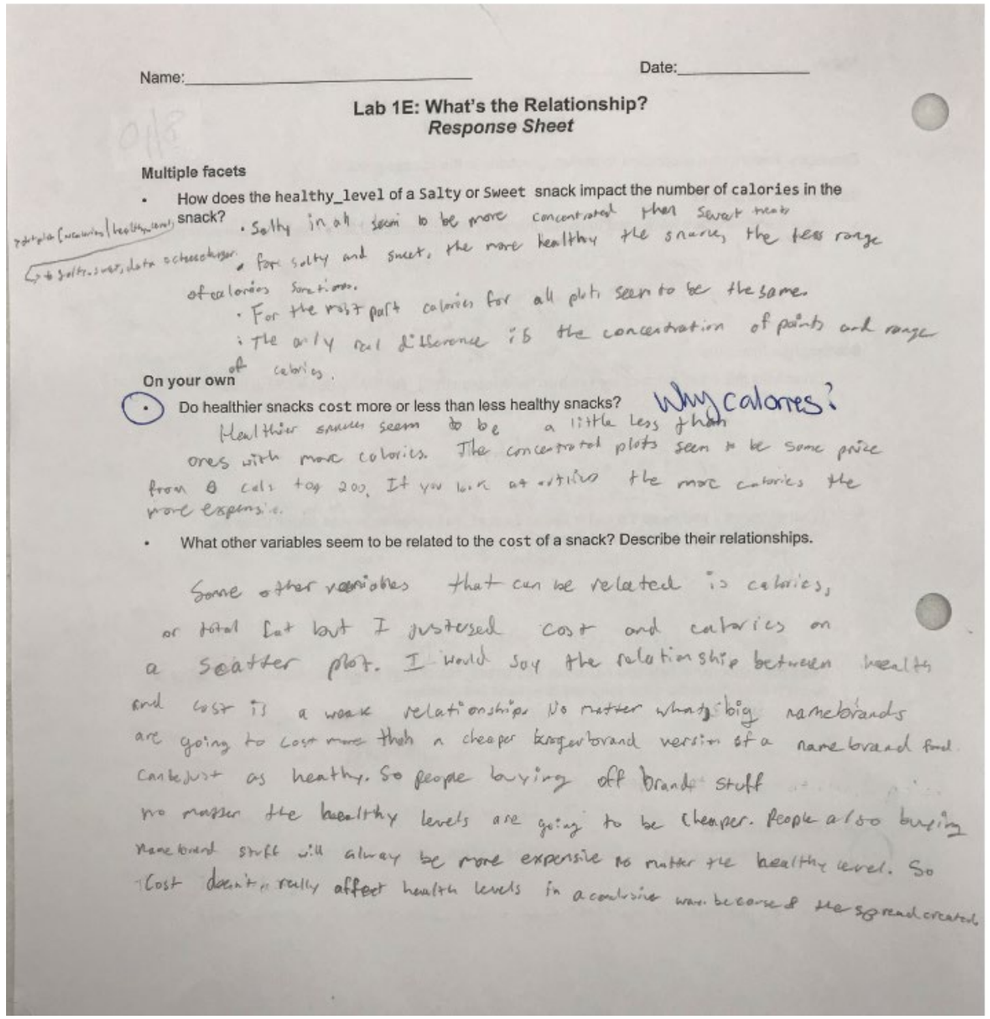
- What happens if you swap the calories and total_fat variables in your code? Does the relationship between the variables change? xyplot(total_fat~calories, data=food)
  The relationship between the variables don't change. Also you can see most of the snacks have low amounts of total fat.

- Does the relationship between calories and total_fat change when the snack is either Salty or Sweet? Write down the code you used to answer this question.
  xyplot(total_fat~calories|salty_sweet, data=food)
  No, the relationship between calories still doesn't change.

4-variable scatterplots

- Create a scatterplot that uses these 4 variables: sodium, sugar, healthy_level, salty_sweet.

xyplot(sodium~sugar|salty_sweet, data=food, groups=healthy_level)



Name:_____ Date:_____

Lab 1E: What's the Relationship?
Response Sheet

Multiple facets

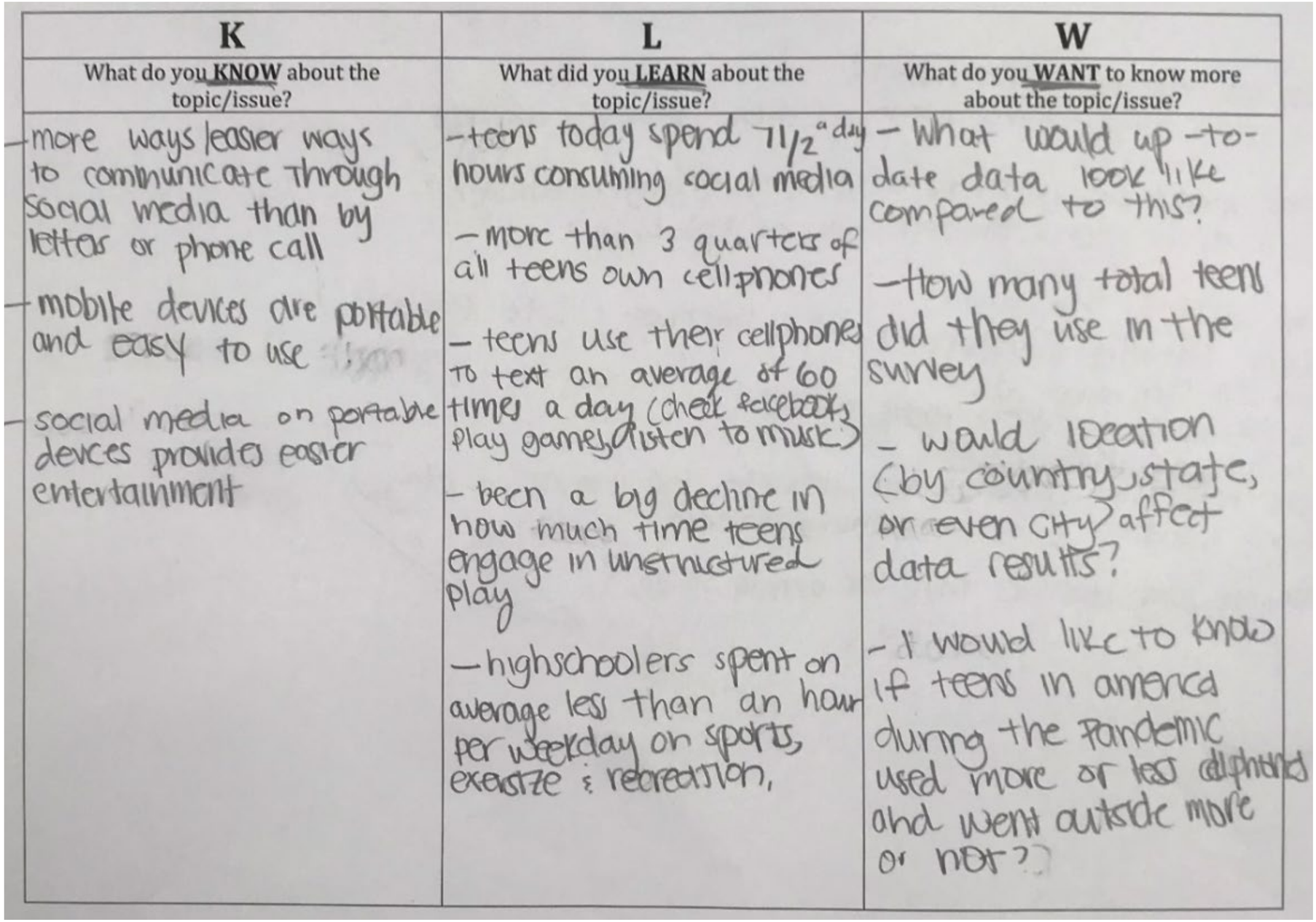
- How does the healthy_level of a Salty or Sweet snack impact the number of calories in the snack?
  - Salty in all seem to be more concentrated than sweet treats for salty and sweet, the more healthy the snacks, the less range of calories sometimes.
  - For the most part calories for all plots seem to be the same.
  - The only real difference is the concentration of points and range of calories.

On your own

- Do healthier snacks cost more or less than less healthy snacks? Why calories?
  Healthier snacks seem to be a little less than ones with more calories. The concentrated plots seem to be some price from 0 cals to 200. If you look at outliers the more calories the more expensive.

- What other variables seem to be related to the cost of a snack? Describe their relationships.
  Some other variables that can be related is calories, or total fat but I justersed cost and calories on a scatter plot. I would say the relationship between health and cost is a weak relationship No matter what big name brands are going to cost more than a cheaper knockoff brand version of a name brand food can be just as healthy. So people buying off brand stuff no matter the healthy levels are going to be cheaper. People also buying name brand stuff will always be more expensive no matter the healthy level. So cost doesn't really affect health levels in a contrive way because of the spread created.

Student Work



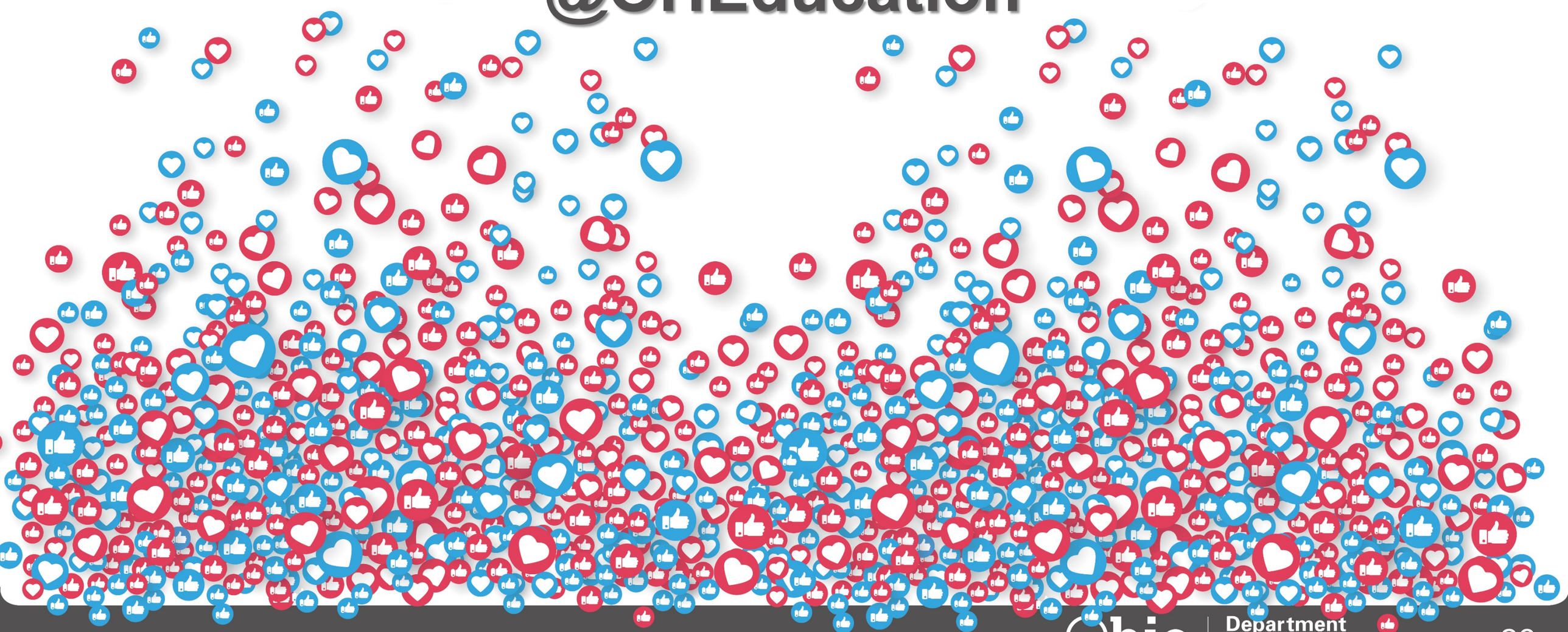
K What do you **KNOW** about the topic/issue?	L What did you **LEARN** about the topic/issue?	W What do you **WANT** to know more about the topic/issue?
- more ways/easier ways to communicate through social media than by letters or phone call - mobile devices are portable and easy to use - social media on portable devices provides easier entertainment	- teens today spend 7 1/2 a day hours consuming social media - more than 3 quarters of all teens own cellphones - teens use their cellphones to text an average of 60 times a day (check facebook, play games, listen to music) - been a big decline in how much time teens engage in unstructured play - highschoolers spent on average less than an hour per weekday on sports, exercize & recreation,	- what would up-to-date data look like compared to this? - how many total teens did they use in the survey - would location (by country, state, or even city) affect data results? - I would like to know if teens in america during the pandemic used more or less cellphones and went outside more or not?

Where to Apply for the Data Science Foundations Pilot?

https://education.ohio.gov/Topics/Learning-in-Ohio/Mathematics/Resources-for-Mathematics/Math-Pathways/Data-Science-Foundations

Q & A

@OHEducation



Share your learning community with us!

#MyOhioClassroom

Celebrate educators!

#OhioLovesTeachers