

# Data Ethics & Data Collection

---

Cara Marta Messina and Laura Johnson  
ENGW 1110 Introduction to First Year Writing  
Kelly Garneau  
Fall 2019



Northeastern University  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point  
during the presentation!*

# Getting to Know You!

What are some recent advertisements you have seen pop up on your Snapchat, Instagram, Facebook, Youtube, and other online spaces?

- Which advertisements interested you?
- What are some surprising advertisements?
- Why do you think you received those advertisements?



# Workshop Agenda

- Objectives
- Introduce 'Big Data' Concepts
- Discuss data, privacy, and data categorization
- Data Creation and Collection: Introducing Google Sheets

Slides, handouts, and data available at <http://bit.ly/2KsBbv0>



# Workshop Objectives

- Understand the ways data is being used in society as well as how algorithms impact and shape our daily lives
- Understand the ways in which technology reflects cultural, social, and political biases.
- Explore the ways in which privacy and security are being reshaped and redefined through the use of big data and algorithms
- Explore how to create, collect, and analyze data in Google Sheets



# What is 'Big Data'?

Big data has been called the 'new oil' by some. Shoshana Zuboff argues that we now live in an era of 'surveillance capitalism,' in which large amounts of information are being analyzed quickly and typically used for profit.

The four components of big data are: **volume**, **variety**, **velocity** and **veracity**



# Big Data: What is it and why should we care?

- Big data is characterized by its **scale**
- Big data **sources** include: digitized records, social media/internet activity, or sensors from the physical environment.
- Big data is often **privately owned**
  - Example: an insurance company purchasing social media activity from facebook in order to make specific insurance sales decisions.



# Facebook Preferences

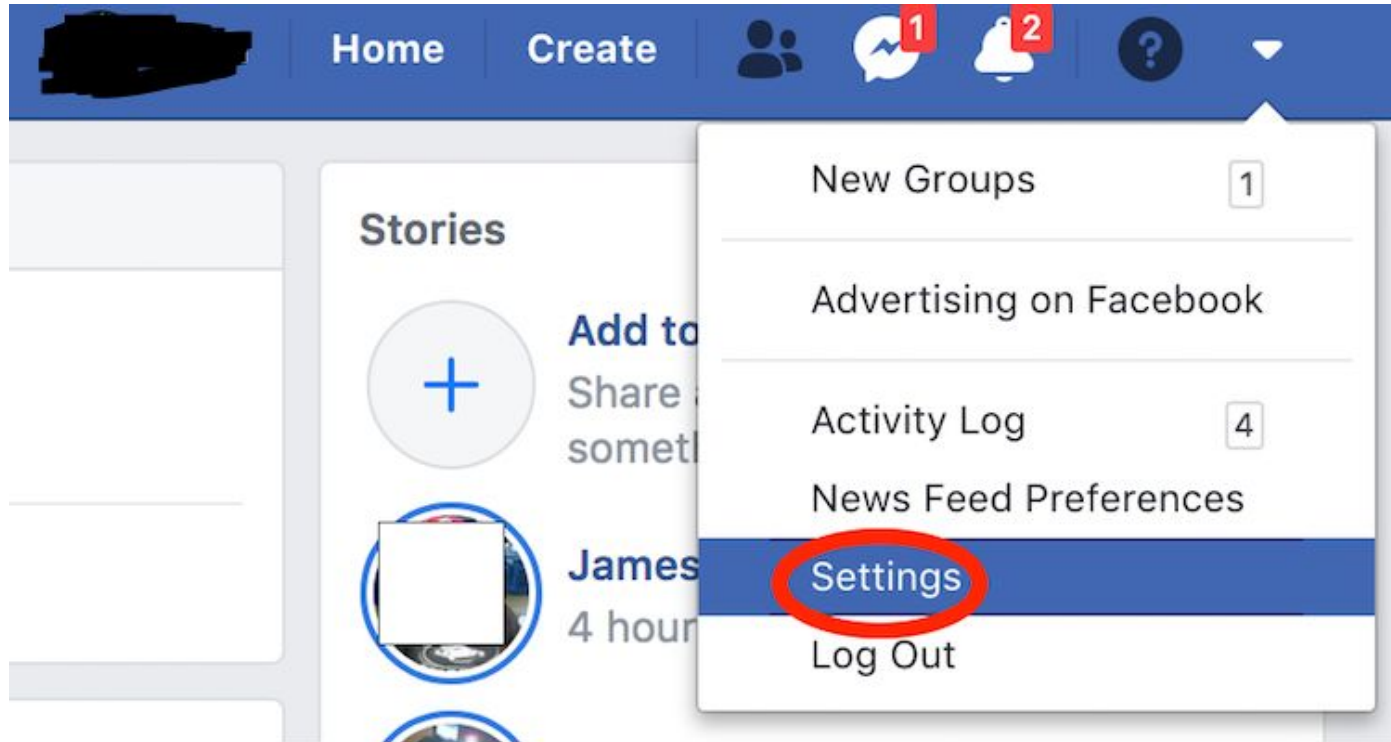
Facebook collects, stores, and sells information about you so you get better targeted ads and your newsfeed is tailored to your categories.

Other social media sites that do this:

- Instagram (owned by Facebook)
- Google
- YouTube (owned by Google)
- Twitter

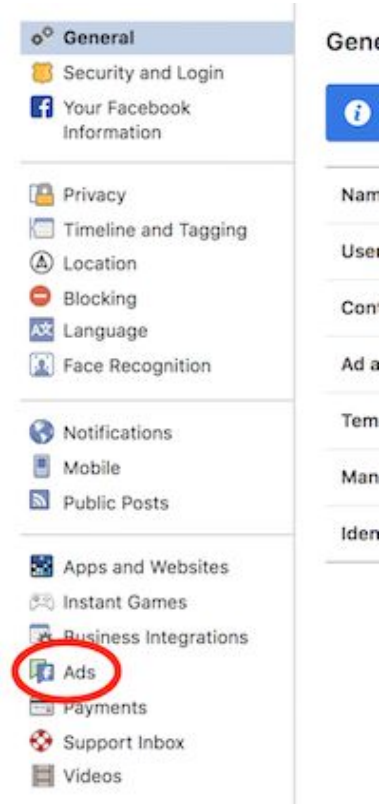


# Settings > Ads > Your information > Categories

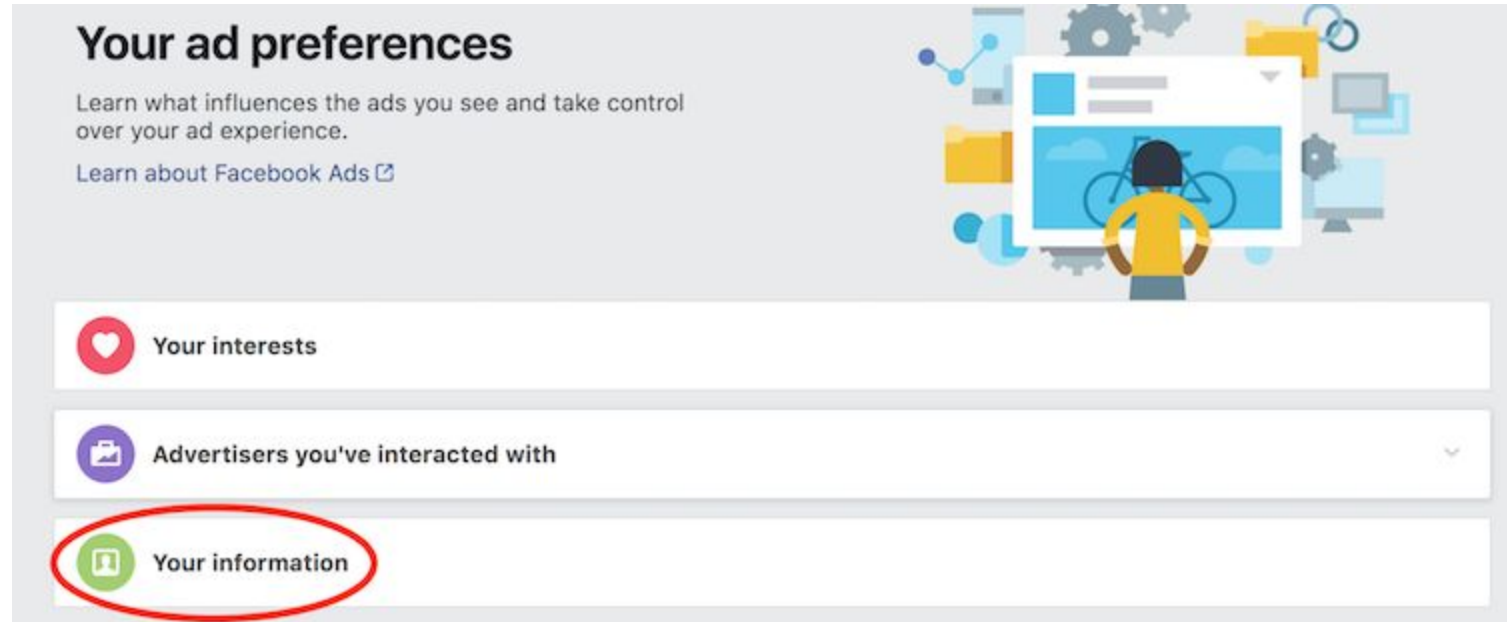




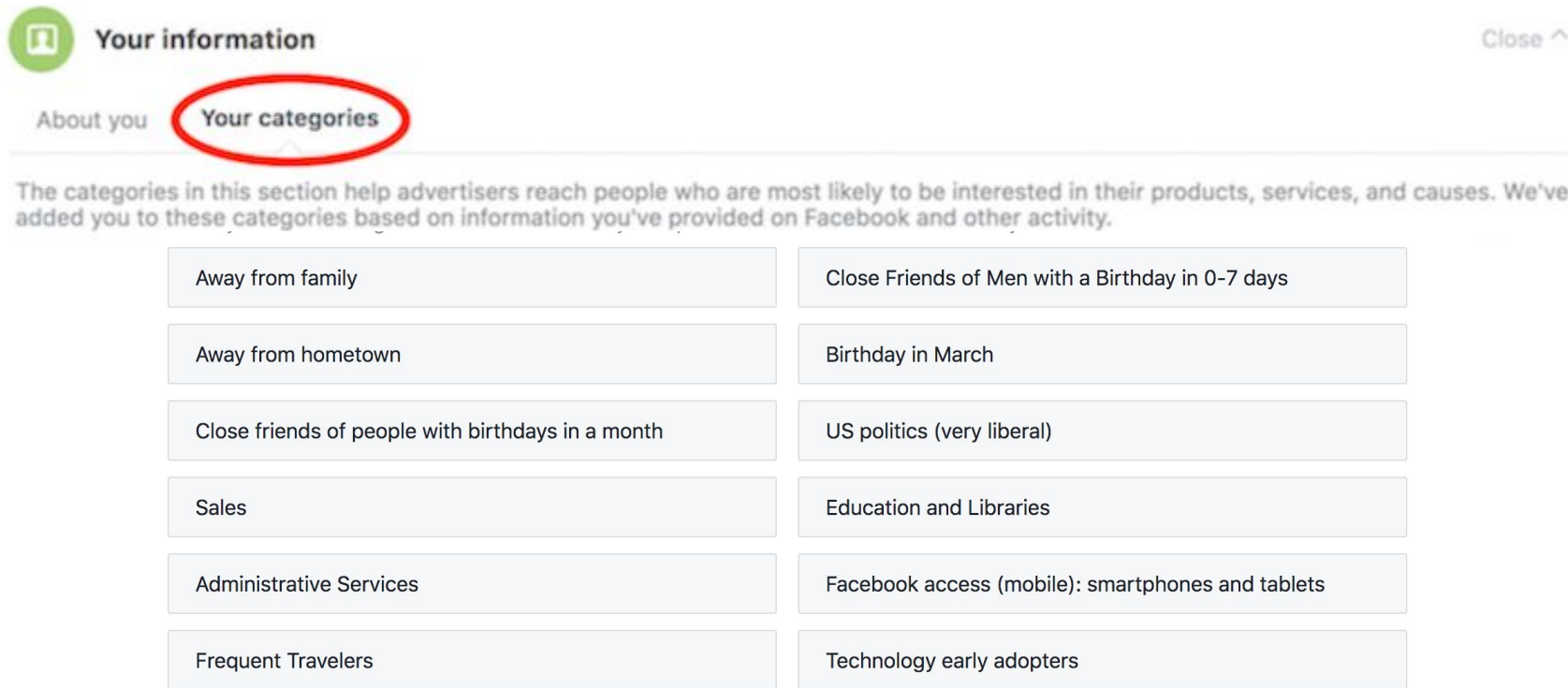
# Settings > Ads > Your information > Categories



# Settings > Ads > Your information > Categories



# Settings > Ads > Your information > Categories



The screenshot shows the Facebook 'Your categories' settings page. At the top, there's a green circular icon with a white 'i' and the text 'Your information'. Below it, two tabs are visible: 'About you' and 'Your categories', with the latter being circled in red. A 'Close ^' link is in the top right. A descriptive paragraph states: 'The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.' Below this, there are two columns of category boxes. The left column includes: 'Away from family', 'Away from hometown', 'Close friends of people with birthdays in a month', 'Sales', 'Administrative Services', and 'Frequent Travelers'. The right column includes: 'Close Friends of Men with a Birthday in 0-7 days', 'Birthday in March', 'US politics (very liberal)', 'Education and Libraries', 'Facebook access (mobile): smartphones and tablets', and 'Technology early adopters'.

**Your information** Close ^

About you **Your categories**

The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.

Away from family	Close Friends of Men with a Birthday in 0-7 days
Away from hometown	Birthday in March
Close friends of people with birthdays in a month	US politics (very liberal)
Sales	Education and Libraries
Administrative Services	Facebook access (mobile): smartphones and tablets
Frequent Travelers	Technology early adopters



# Google's File on You is 10 Times Bigger Than Facebook's — Here's How to View It

Google, Amazon, Apple, and Microsoft are all central players in “surveillance capitalism” and prey on our data.



Example: If you have **location services** turned on for Google (if you use Google Maps), Google can track your every move. Go to:

<https://www.google.com/maps/timeline>



# Ethical Implications

- Cambridge Analytica Controversy
- Big data also raises questions of autonomy, anonymity, privacy, discrimination, and bias.
- Questions to consider:
  - How are we being represented online?
  - How is our data being used?
  - Who is using it and for what purposes?
  - How might it be used in the future?



# DIY Cybersecurity and Tightening your Privacy

Want to make your life more private? Follow this “DIY Guide to Feminist Cybersecurity”

<https://hackblossom.org/cybersecurity/>



# Algorithms and Bias



Northeastern University  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# Activity: Adopt or Not?

**Small Group:** Find a partner or two! You all work for an adoption agency and have to decide if someone can adopt a dog. On your handouts, please read the four previous adoption applications and decide if the new adoption applicant can adopt or not.

**Do you think this new applicant can adopt a dog? Why or why not?**





# Class Discussion: Adopt or Not?

- Do you ACCEPT or REJECT their application? Why?
- What questions from the application did you weigh more? Why?
- What might be some implicit biases in this application and in your choices?



# Adopt or Not? Algorithm

Algorithms to “read” through applications exist.

- Where might you see these algorithms being used to make decisions?
- What biases may be ingrained in those applications?
- How might the impact choices?



# So what does 'big data' have to do with First Year Writing?



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# Documenting Your Writing

For your upcoming assignment, you will be collecting your own data by tracing your writing practices using **Google Sheets**.

You will create your own spreadsheet with columns and collect data for a week!



# Questions to Consider

As we talk about Google Sheets and what it can do, think about:

- In what ways are your column choices impacting how you view your writing patterns?
- Which information will you choose to omit or highlight, both as you are collecting data and then as you're reflecting on that data?
- How do you think about your audience and your purpose while you're writing? How can you highlight this through your data collection?



# Google Sheets

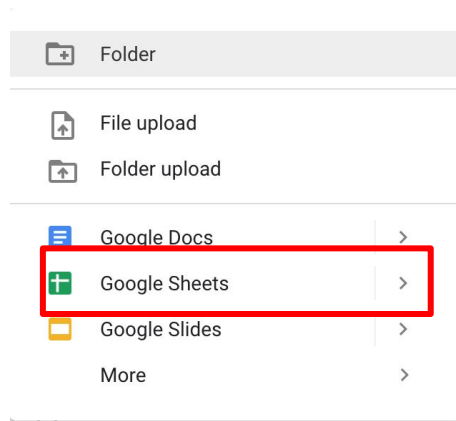
**Google Sheets** is Google's spreadsheet software; it's like Excel. You can create spreadsheets and use functions to collect, organize, analyze, and interpret data (both numerical and textual data).

Feel free to download the Example Sheet from our GitHub and open it with Google Sheets to follow along.



# Creating A Google Sheet

- Go to your Google Drive (drive.google.com)
- Folder Creation:
  - If you have a lot of files/folders, choose the folder that this sheet would suit best
  - Otherwise, create a new folder and title it clearly
- Create a Sheet



# Google Sheets Anatomy

Title of  
the file

Share function  
to add Kelly



Example Sheet



Toolbar

File Edit View Insert Format Data Tools Add-ons Help [All changes saved in Drive](#)



Share

Columns  
(letters)

Rows  
(numbers)

Cell (B2)

Sheet1

To add more sheets in one file





# Important Google Sheets Vocabulary

**Rows:** the vertical numbers

**Columns:** the horizontal letters

**Cells:** the individual rectangular box that has a unique ID based on placement (column + row)

**Charts:** visualizations that can be created about your data

**Functions:** mathematical functions that can be used on numerical data

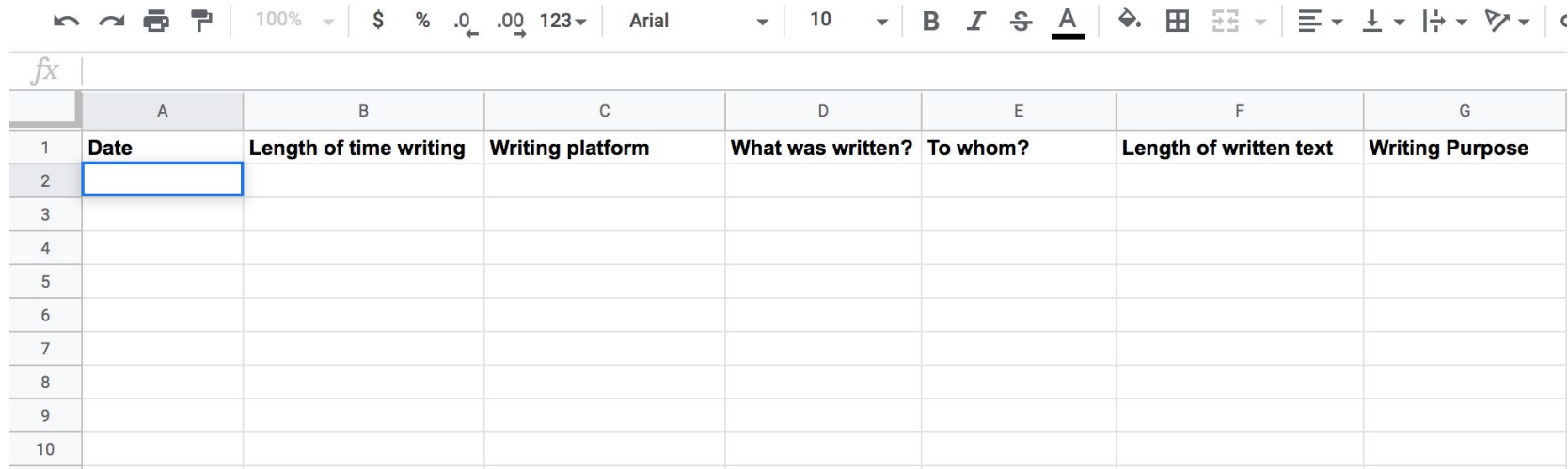
**Data Validation** the input specifications for particular cells

**Conditional Formatting** filter for data that lets you highlight or mark information (blank vs not blank)



# Organizing Your Writing Data

Create several columns to enter in data about your writing patterns. For example:



The image shows a screenshot of a spreadsheet application. At the top is a toolbar with various icons for undo, redo, copy, paste, and formatting. Below the toolbar is a formula bar with the text "fx". The spreadsheet itself has a grid with columns labeled A through G and rows numbered 1 through 10. The first row (row 1) contains the following headers: "Date", "Length of time writing", "Writing platform", "What was written?", "To whom?", "Length of written text", and "Writing Purpose". The second row (row 2) has a blue border around the cell in column A, which is currently empty. The remaining rows (3 through 10) are also empty.

	A	B	C	D	E	F	G
1	<b>Date</b>	<b>Length of time writing</b>	<b>Writing platform</b>	<b>What was written?</b>	<b>To whom?</b>	<b>Length of written text</b>	<b>Writing Purpose</b>
2							
3							
4							
5							
6							
7							
8							
9							
10							



# Example - available on <http://bit.ly/2KsBbv0>

A	B	C	D	E	F
Date	Length of time writing (mins)	Writing platform	To whom?	Length of written text (characters)	Writing Purpose
09/25	5	social media	Friend	250	Vent
09/25	10	email	Teacher	300	Extension
09/25	11	email	Family	230	Say hi
09/25	4	texting	Family	20	Check in
09/26	6	forums	Internet Stranger	560	Persuade
09/26	7	social media	Friend	350	Jokes
09/27	15	forums	Internet Stranger	700	Persuade
09/27	23	word processing softw	Teacher	300	Get Good Grade
09/27	7	social media	Friends	120	Vent
09/27	2	texting	Family	30	Check In
09/27	2	texting	Friend	90	Jokes
09/27	2	texting	Friend	30	Making plans
09/28	90	word processing softw	Teacher	350	Get Good Grade
09/28	2	texting	Friend	150	Making plans
09/29	30	word processing softw	Teacher	300	Get Good Grade
09/29	32	word processing softw	Teacher	200	Get Good Grade
09/29	10	forums	Internet Stranger	500	Persuade
09/29	11	email	Teacher	100	Question about work



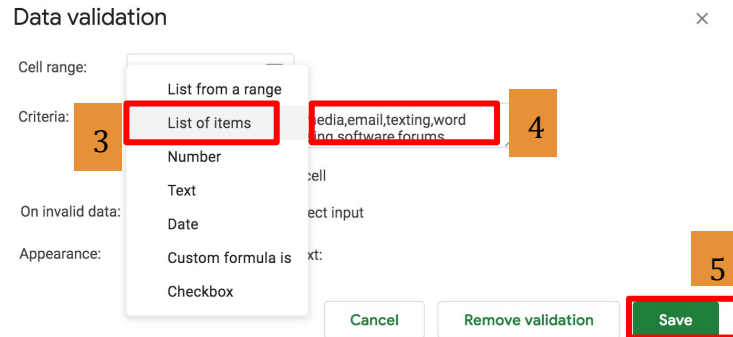
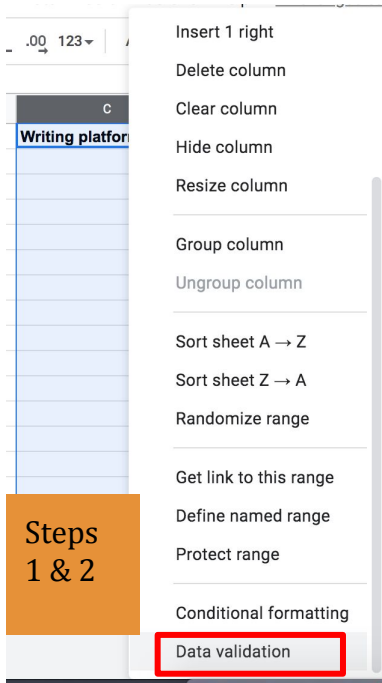
# Data Validation

Some columns might require “data validation,” which means only particular data can be selected.

[illegible]

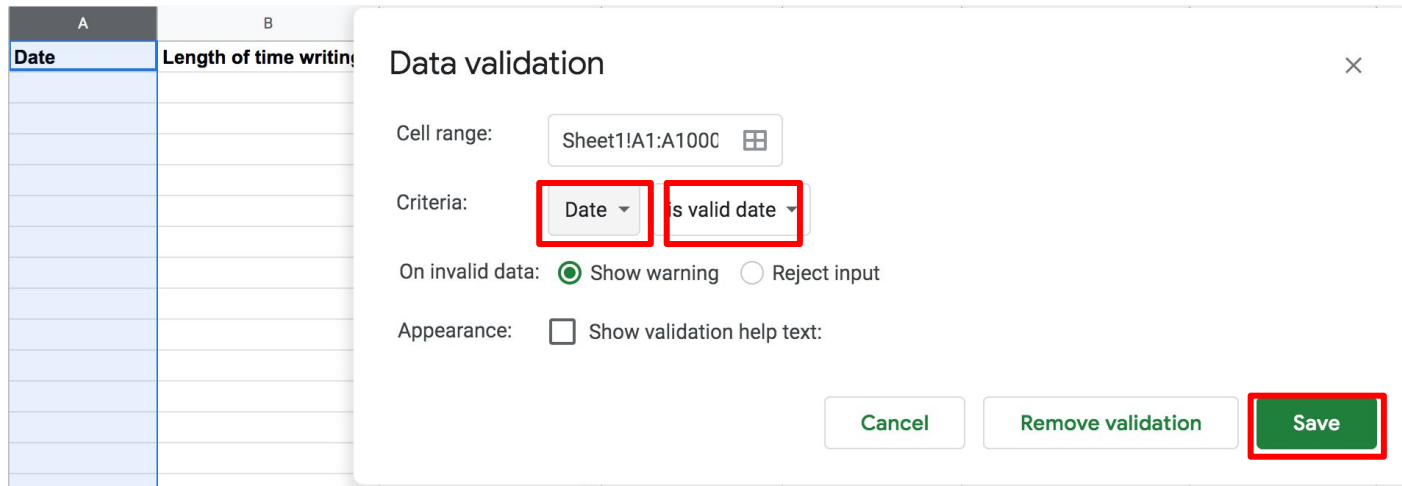
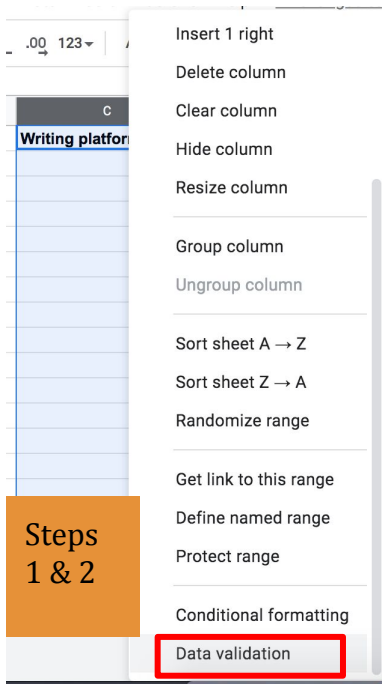
# Setting Up Data Validation

1. Select the section you want to be validated (for example, a column)
2. Right click and scroll to “Data Validation” (or go to Data > Data Validation)
3. Choose the proper criteria (in this case, “List of Items”)
4. Input the list of items separating each item with a comma
5. Click “Save”



# Date Data Validation

1. Follow the same steps as 1 & 2
2. For criteria, click “Date” and then click “Is a Valid Date”
3. Now you can only enter date information in that cell!



# Functions

When you're finally done collecting all your writing data, what do you do with it? You can use **functions** to summarize your data!

Example:

- Average number of characters written per entry
- Total number of characters written
- Total length of time writing
- Average length of time writing



# How to do Functions

1. In a blank cell (preferably at the bottom of your data), type =
2. You will be provided options for different functions!
  - a. To do a total, type =SUM(
  - b. To do an average, type =AVERAGE(
3. Select the cells you would like to be included in the total
4. Voila! You have your function that will also update if you update cells!

11	
4	
6	
7	
15	
23	
7	
2	
2	
2	
2	
90	
2	
30	
32	
10	
11	

B	
	23
	7
	2
	2
	2
	90
	2
	30
	32
	10
	11

269 ×
=SUM(B2:B19)

14.94444444 ×
=AVERAGE(B2:B19)

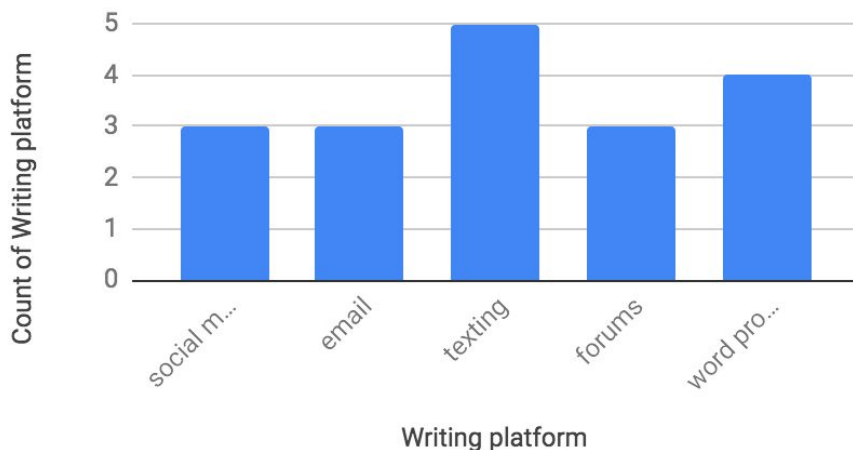




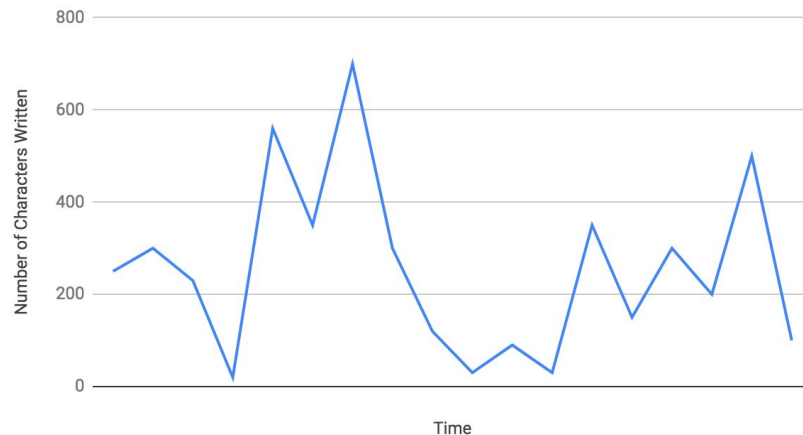
# Charts

Similar to functions, charts allow you to summarize your information as well as visually present it. You can also use charts for your “data validation” sections!

Count of Writing platform

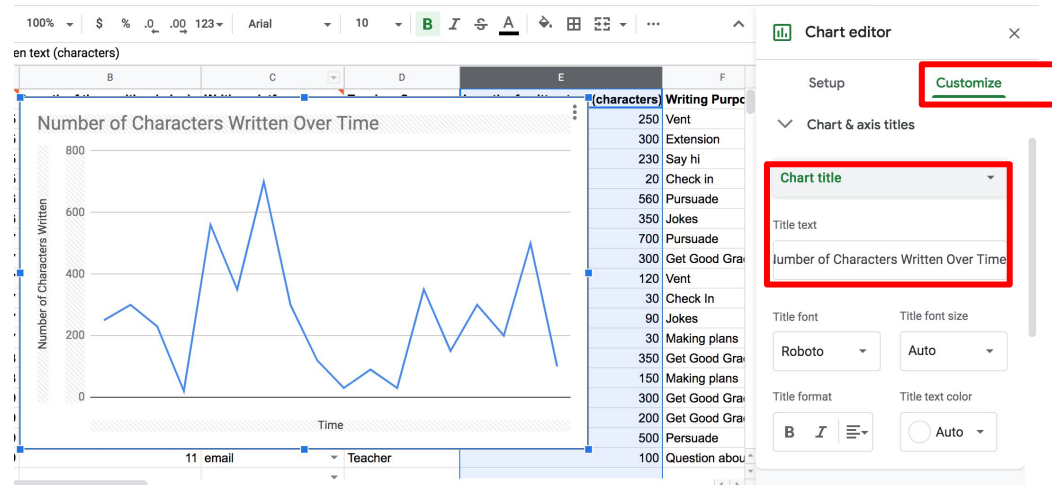
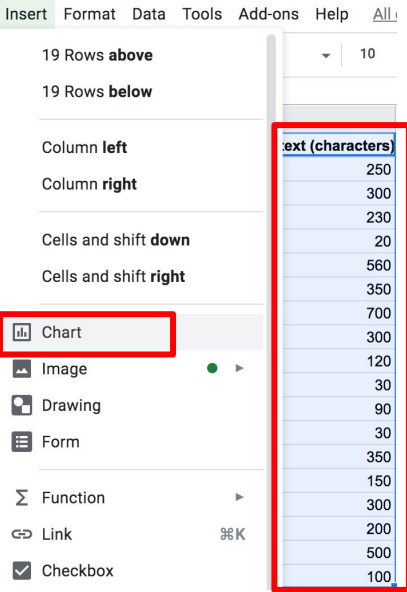


Number of Characters Written Over Time



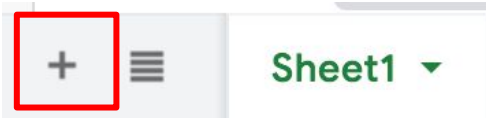
# How to Add a Chart

1. Select the data you want to chart
2. Click “Insert” then “Chart”
3. Customize your new chart using the “Chart Editor”
4. Add labels (the title, x and y axis, etc) in “Customize” > “Chart & Axis Title”



# Your Turn!

Make a **second sheet** (click the + sign in the bottom) to start experimenting with Google Sheets:



1. Create at least five columns and add at least 5 rows of data in for each
2. Use the “Data Validation” on one column
3. Find the SUM and AVERAGE of one of your columns
4. Create one chart

Reminder: these slides are available at <http://bit.ly/2KsBbv0>



# Post-Exploration Group Discussion

1. What type of data did you create?
2. What did your chart look like?
3. What did you include in your data validation?
4. Anything surprising? Any questions?



# Thank you!

If you have any questions, contact us at:

**Cara Marta Messina**

Digital Teaching Integration

Assistant Director

[messina.c@husky.neu.edu](mailto:messina.c@husky.neu.edu)

**Laura Johnson**

Digital Teaching Integration

NULab Coordinator

[johnson.lau@husky.neu.edu](mailto:johnson.lau@husky.neu.edu)

Slides and handouts available at <http://bit.ly/2KsBbv0>

Office Hours: **Tuesdays from 1–3PM in 401 Nightingale Hall**



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point  
during the presentation!*