WELCOME TO GENERAL ASSEMBLY

WIFI
GA Guest

PASSWORD
yellowpencil

SCHEDULE
5:30: Workshop
7:15: Q&A


GA
GENERAL ASSEMBLY

OUR COURSES - Data Analytics

	Data Analytics Part Time	Data Science Part Time
Length	10 weeks, part time - 2 nights/week	10 weeks, part time - 2 nights/week
Next Lesson	Aug 29	Sep 11
Hours	~60 hours	~60 hours
Outcome	Learn Excel, SQL, and Tableau	Learn Python, predictive modeling, and scripts
Tuition	$3,950 (payment plans)	$3,950 (payment plans)

MICHAEL VALERI

Sr. Game Data Analyst, Big Fish Games

ALSO...

LET'S GET STARTED...

OPENING

LEARNING OBJECTIVES

Part 1 - Introduction (30 min) - 6:00 MAX

- Explain the relevance of data analysis across industries and job functions.
- Explain how analysts “think” about data. **The WORKFLOW OF AN ANALYST or WFOAA**
- Explain tools analysts use to tell a **story** with data.

Part 2 - Hands on Learning (1 hr 15 min) - 7:15 MAX

- Evaluate the quality and structure of a dataset, using the **WFOAA.**
- Use Google Sheets to perform descriptive and exploratory analysis on datasets, using the **WFOAA.**
- Use data analytics to tell a **story** and inform business decisions using the **WFOAA.**

Part 3 - Ask Me Anything - 7:30 MAX

DATA ANALYTICS 101

PRE-WORK

PRE-WORK

- Are you able to access Google Sheets?
  - If yes, awesome!
  - If not, create a free Google account by visiting: accounts.google.com

DATA ANALYTICS 101

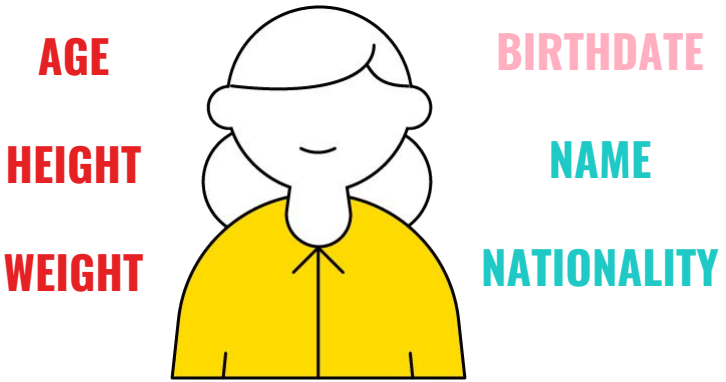
INTRODUCTION: DEFINING DATA

DEFINING DATA

- We live in a world filled with data, it's all around us. But, if it's all around us, what is it?
- Data is simply *information* that exists in a variety of formats and sizes.
- Data can be thought of as an "object" but it's not necessarily seen at first glance.
- Data is POWERFUL → TARGET STORY
- You've created data in many ways today, perhaps by:
  - Logging into a website
  - Making a purchase with a credit/debit card
  - Taking public transportation

DEFINING DATA - Fitbit EXAMPLE

- This data, only about you, represents a variety of data types.
- Your age, height, and weight are all numbers.
- Your birthday is a date.
- Your name and nationality are text (or strings, as we call them in data analytics).



DEFINING DATA

- Not only do you have data attached to you, but you're creating data with every turn.

- Say hello to your neighbor!
- Work with your neighbor to generate a list of 3-5 ways you have you created data today.

DEFINING DATA - WHAT IS DATA ANALYTICS

- Data Analytics is the process of transforming raw data, perhaps like the types you just mentioned, with the purpose of drawing conclusions about that information.
- Or more plainly….**STORYTELLING with NUMBERS.**

DATA ANALYTICS 101

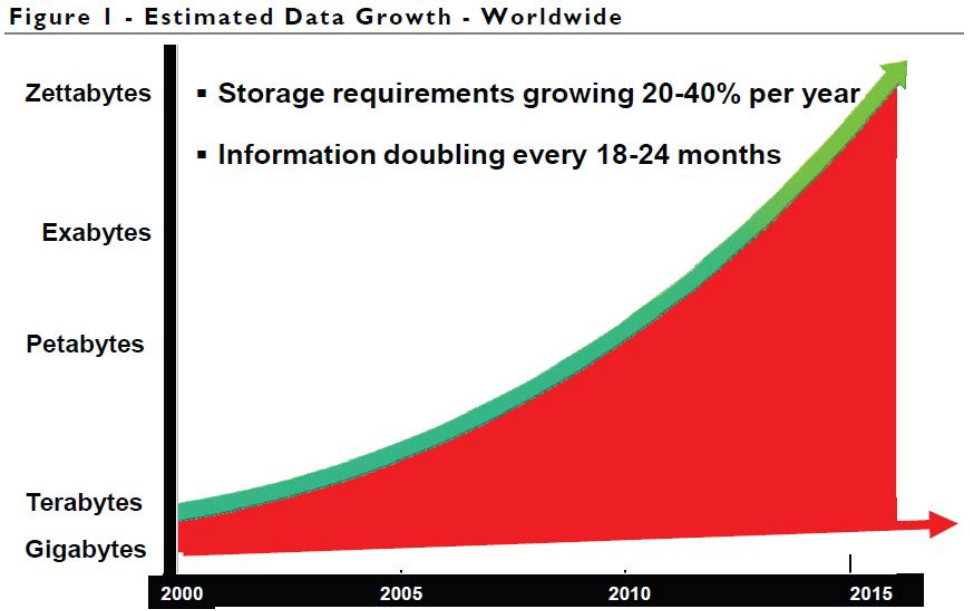
LESSON: HOW DATA ANALYSTS THINK ABOUT DATA

HOW DATA ANALYSTS THINK ABOUT DATA - THERE IS ALOT

- The rate at which data is being created is rapidly accelerating. The key for successful data analysis is taking data and forming actionable conclusions and insights. ZETABTYE = 1 SEXTILLION BYTES. **GOLD RUSH ON DATA**

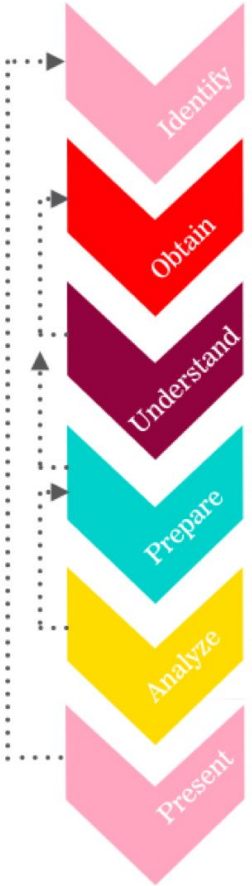
- Data Analysts do this by using a “workflow” to guide to them through the process.

- What might be some of the steps in this process?



HOW DATA ANALYSTS THINK ABOUT DATA

- Identify the Problem
  - Before you begin working with any data, you must understand the problem that you're trying to answer. **Eliminate ambiguity**.
- Example from Business Partner at Amazon: "I want to know how **large** our Spanish eBook Business is in the Kindle Store."



IDENTIFY THE PROBLEM

OBTAIN THE DATA

UNDERSTAND THE DATA

PREPARE THE DATA

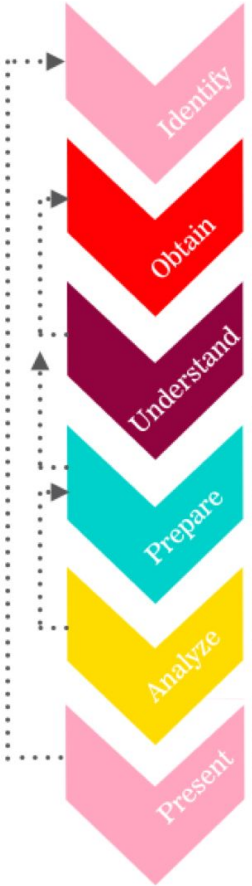
ANALYZE THE DATA

PRESENT THE RESULTS

HOW DATA ANALYSTS THINK ABOUT DATA

- Obtain the Data
  - To work with the data, you first have to find it or collect it, and it has to be the right data to help you answer the question.

- Example from Business Partner at Amazon: Find where data lives, write SQL query, execute data warehouse job.



IDENTIFY THE PROBLEM

OBTAIN THE DATA

UNDERSTAND THE DATA

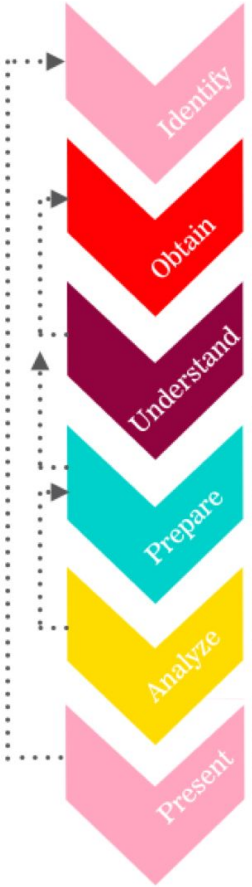
PREPARE THE DATA

ANALYZE THE DATA

PRESENT THE RESULTS

HOW DATA ANALYSTS THINK ABOUT DATA

- Understand the Data
  - Then you need to ensure you can correctly interpret the results and trust the data.
- Example from Business Partner at Amazon: Look at columns obtained, row count, etc. See if there multiple currencies for $, all in USD, etc.



IDENTIFY THE PROBLEM

OBTAIN THE DATA

UNDERSTAND THE DATA

PREPARE THE DATA

ANALYZE THE DATA

PRESENT THE RESULTS

HOW DATA ANALYSTS THINK ABOUT DATA

- Prepare the Data
  - Prep your workbook
    - Freeze panes
    - Zoom out and increase text size
    - Etc.
  - You should making sure the data doesn't contain incorrect or missing values.

HOW DATA ANALYSTS THINK ABOUT DATA

- Analyze the Data
  - Now, you are ready to uncover the answer to your question, assuming you haven't ended up at a prior step due to missing data or a poorly understood question.

HOW DATA ANALYSTS THINK ABOUT DATA

- Present the Results
  - Assuming you find what you are looking for, and it seems compelling enough to share with others, you need to determine the best way to share your results with others.

TELL A STORY WITH CHARTS, TABLES, INSIGHT, ETC. PROVIDE ADDITIONAL CONTEXT, GIVE MORE THAN THEY ASKED FOR!

HOW DATA ANALYSTS THINK ABOUT DATA - MOST IMPORTANT

- These general steps are necessary for each and every data analysis you do; but like a snowflake, each time will be a little different, as well.

- One of the most important things you'll notice is that the workflow is not strictly linear. Even though you begin at the top and end at the bottom, you will revisit various steps along the way as needed.

- Instead of thinking of this as step-by-step instructions for doing data analysis, you should think about these stages as the guiding principles for analyzing data. Experience in navigating these competing factors is what separates a good data analyst from a great one.

DATA ANALYTICS 101

LESSON: TOOLS OF THE DATA ANALYST

TOOLS OF THE DATA ANALYST

- In order to proceed through the workflow, a data analyst uses a suite of tools to assist along the way.

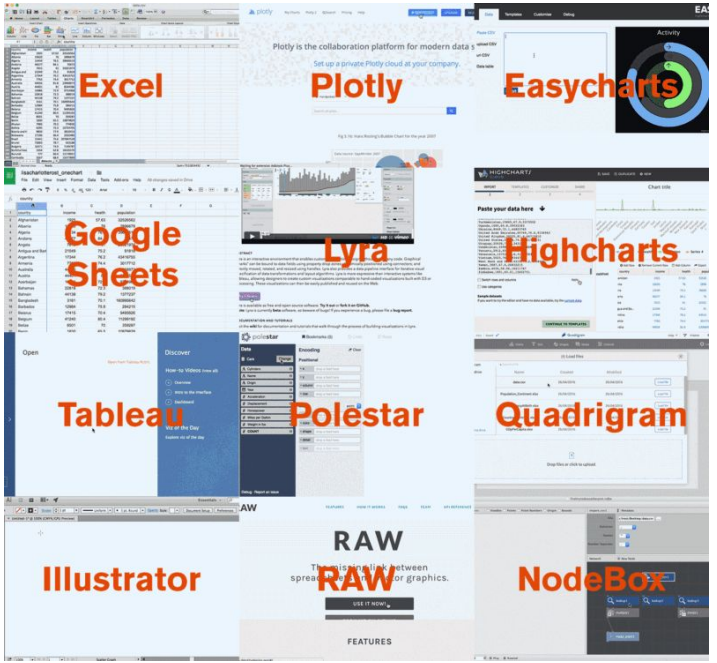
- While data analysts can apply a specialized suite of tools, the analyst's judgement and intuition is the most important tool.
  - Is there a tool that will "Identify the Question" or "Understand the Data" for you? Likely not; the analyst themselves become the best tool in this scenario.

- A Data Analyst may use tools like: Google Sheets, Excel, SQL, or Tableau.

TOOLS OF THE DATA ANALYST

- One thing is for sure: there is no shortage of data tools to use for purposes that range from analysis to visualization.

- TWO CATEGORIES
  - MINING
    - SQL → Language to pull data
    - R → Language to predict future behavior
  - REFINING
    - Tableau, Excel, Google Sheets, Excel



DATA ANALYTICS 101

On Hands Learning: USING DATA TO INFORM BUSINESS DECISIONS

USING DATA TO INFORM BUSINESS DECISIONS

- It's time to get our hands dirty working with real data. It looks like a job for you:

  Chipotle Founder, Steve Ells, has hired a team of General Assembly (GA) Analysts to help him assess the current state of Chipotle to help plan for future growth of the business. Steve has task the GA Analysts with helping him document Chipotle's current growth patterns across geographies and to better understand the "sides" purchased by customers at Chipotle restaurants.

  The GA Analysts will provide insights for Chipotle's strategy related to:
  (1) restaurant expansion - Specific? Get better ?'s
  (2) menu changes - Specific?

- Before we jump into any analysis, let's watch a clip to better understand Chipotle's business.

!REMEMBER FROM THE WFOAA (Workflow Of An Analyst)
STEP 1: IDENTIFY THE PROBLEM / TOOL: YOU (MOST IMPORTANT)

USING DATA TO INFORM BUSINESS DECISIONS

- We will want to consider Chipotle's strategy in two directions: expansion plans and menu alterations.
- We have a clear direction on which "business" questions to focus our analysis on, but we will want to provide insights based on "data" questions.

- Working in pairs, consider what types of data we might need to gather to answer these questions.
  - Consider the "information" we would want to gather.
  - What sources would we gather this from?

DATA ANALYTICS 101

GUIDED PRACTICE: ANALYZING PROPERTY GROWTH AT CHIPOTLE

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- Let's access the Google Sheet we will be using for the remainder of today.
  - Visit this link: goo.gl/gnEvP6
  - Google sheet functions https://support.google.com/docs/table/25273?hl=en
  - Next, you'll need to click on "File" and "Make a copy…" to create a version of the file on your own Google Drive.
    - This Google Sheet is in "View only" mode, so you can't edit it.
  - When you copy the file to your Google Drive, you'll have full edit access to the copied version!

!REMEMBER FROM THE WFOAA!
STEP 2: OBTAINING THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

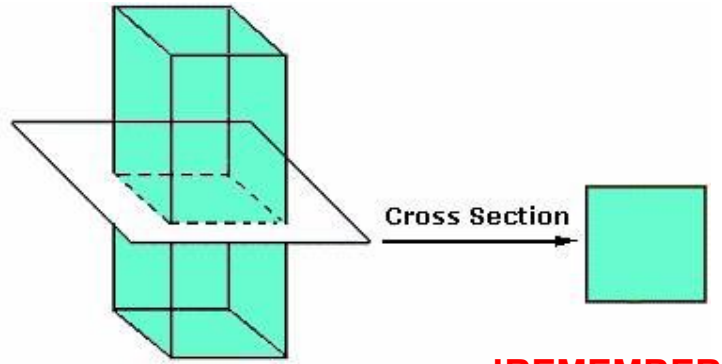
- Let's talk through some important vocabulary related to the Google Sheets environment:
  - **Workbook:** A collection of worksheets.
  - **Worksheet:** The area where data is arranged and calculations are performed.
  - **Column:** A vertical collection of cells.
  - **Row:** A horizontal collection of cells.
  - **Cell:** The intersection of a column and row on a worksheet.
  - **Array:** A collection of cells in a row, column, or across rows or columns.
  - **Function:** Instructions for calculations.

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- We are getting ready to start **exploratory data analysis** (EDA), where we summarize and display data in such a way that interesting features become apparent **(Starts to tell a story)**
  - Data is often “rectangular” (i.e. defined by row and columns); EDA uses graphical and non-graphical methods to break down this “rectangle” into more visually compelling and easy to understand mediums

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- Let's look specifically at the structure of Chipotle's property data in 2015.
  - Data observed in the same time period is known as **cross-sectional data**.



!REMEMBER FROM THE WFOAA!

STEP 3: UNDERSTAND THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- Let's prepare our workbook for analysis
  - Once saved to share drive should be able to freeze the top row
    - Highlight Row 2 go to View ---> Freeze ---> 1 Row

!REMEMBER FROM THE WFOAA!
STEP 4: PREP THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- 2015 Properties
- How many total properties did Chipotle operate?
  - Syntax: =SUM(number1, [number2],…)
  - Solution in J3:=SUM(B2, B3, …, B52) or =SUM(B2:B52)= 2010
- How many Chipotle properties did Chipotle operate in the US?
  - Solution in J4:=SUM(B2:B48) =1987
- How many Chipotle properties did Chipotle outside the US?
  - Solution in J5:=j3 - j4

!REMEMBER FROM THE WFOAA!
STEP 5: ANALYZE THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- 2015 Properties
- Which geography had the most Chipotle properties?
  - Syntax:=MAX(number1, [number2],...)
  - Solution in J7: =MAX(B2:B52)= 351
  - 351 is the largest number of properties in a geography, but does not indicate the geography itself. California has the most Chipotle properties.
- Which geography had the least Chipotle properties?
  - Syntax:=MIN(number1, [number2],...)
  - Solution in J7:=MIN(B2:B52)=1

- Note Sort Max to Small to ensure you capture any multiples
  - Note: There are multiple geographies with one Chipotle property, North Dakota, Vermont, Wyoming, Germany.

!REMEMBER FROM THE WFOAA!

STEP 3: ANALYZE THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- **2009 - 15 Properties**
- Freeze top row
- INSERT ROW ABOVE ROW 1
- Change Years at top to ‘2015, 2014, etc.)
- Calculate growth from 2009 - 2015
  - Since some states have blank or 0 restaurants in 2009 need to use IFERROR
- Column I (=IFERROR((B2-H2)/H2,””)
  - With cell selected double click on box in bottom right corner to auto fill

!REMEMBER FROM THE WFOAA!

STEP 5: ANALYZE THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- **2009 - 15 Properties**
- To build the scatter plot, start by summing the total number of properties for each year to complete the table:

	A	B	C	D	E	F	G	H	I	J	K
1	**TOTALS**	2,010	1,783	1,595	1,450	1,230	1,084	956			
2	**Geography**	**2015**	**2014**	**2013**	**2012**	**2011**	**2010**	**2009**	**% Change from 09 - 15**	**rom Chipotle from 2009**	
3	Alabama	11	9	7	5	3	3	3	267%		
4	Arizona	74	64	63	57	49	41	35	111%	Let's analyze the trend of	
5	Arkansas	5	4	2	2	2					
6	California	351	325	288	234	198	165	142	147%	Year	Total Properties
7	Colorado	76	74	72	71	71	71	70	9%	2015	2010
8	Connecticut	19	14	14	10	7	3			2014	1783
9	Delaware	4	3	3	3	2	1			2013	1595
10	District of Columbia	22	19	17	13	11	8	7	214%	2012	1450
11	Florida	116	99	84	73	65	58	49	137%	2011	1230
12	Georgia	36	32	25	18	16	13	12	200%	2010	1084
13	Idaho	4	3	2	1	1				2009	956

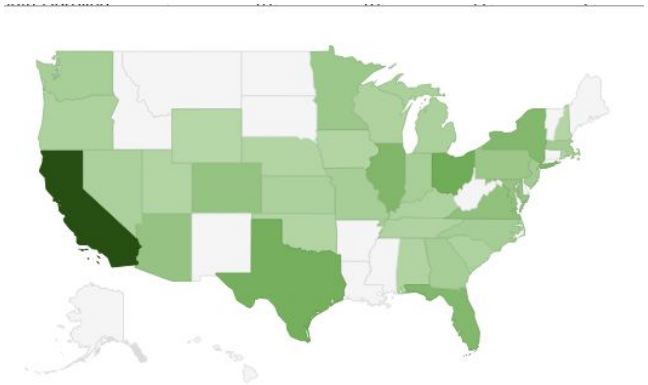
!REMEMBER FROM THE WFOAA!

STEP 5: ANALYZE THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- **2009 - 15 Properties**
- Let's visual # of stores by state

- Sort % change largest to smallest
  - Filter out Blanks and 0
- Highlight States and 2015 # of Stores select Chart Icon scroll down to maps and select Regional
  - Before hitting insert select the Customization Tab and Select United States for Region
  - Select light green for min middle green mid and dark green max

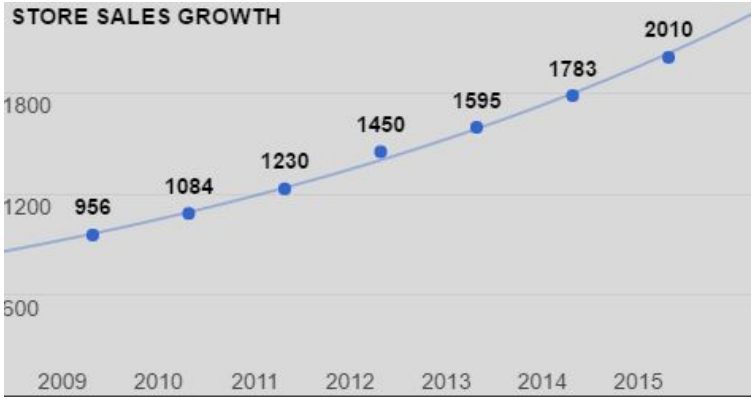


!REMEMBER FROM THE WFOAA!

STEP 5: ANALYZE THE DATA / TOOL: GOOGLE SHEETS

ANALYZING PROPERTY GROWTH AT CHIPOTLE

- **2009 - 15 Properties**
- Let's analyze the trend of overall growth in Chipotle's properties. Create a scatter plot of this data.
  - A **scatter plot** is a diagram that visualizes two variables through the use of coordinates to place each point on a grid.
  - The plot has an **x-axis** (horizontal axis) and **y-axis** (vertical axis).
  - A data point is defined by its value on the x-axis and y-axis in an **ordered pair**: (x,y)
- Edit Scatter plot
  - Select drop down upper right (due to limited time will not actually edit, please go through process after class)



!REMEMBER FROM THE WFOAA!

STEP 5: ANALYZE THE DATA / TOOL: GOOGLE SHEETS

DATA ANALYTICS 101

GUIDED PRACTICE: ANALYZING SALES DATA AT CHIPOTLE

STRUCTURING SALES DATA AT CHIPOTLE

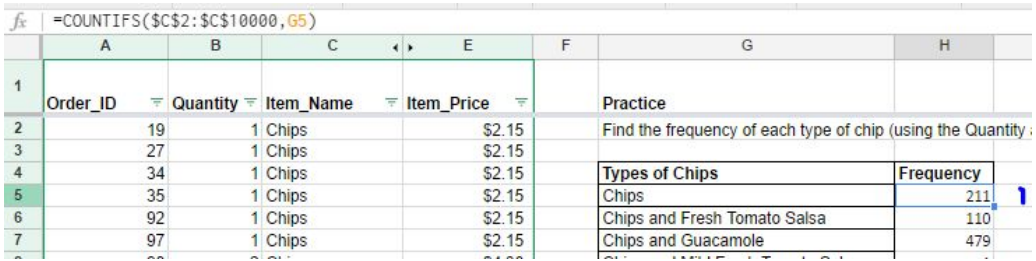
- Sales Data (Chips)
- First, what do we know about this data?
  - The data contains approximately 3,000 meals in about 1,800 Grubhub orders from July to December 2012.
  - Most of the sales data comes from two Chipotle restaurants, one in Washington, D.C., and another in East Lansing, Michigan.
    - Ordering behavior might vary around the country!

!REMEMBER FROM THE WFOAA!

STEP 3: UNDERSTAND THE DATA / TOOL: GOOGLE SHEETS

ANALYZING SALES DATA AT CHIPOTLE

- Sales Data (Chips)
- A **frequency distribution** shows the number of instances a variable takes on a specific value.
  - In our dataset, we want to see how often and in what combination "Chips" are ordered
  - Find the frequency (# of times) of each chip combo was ordered
  - Use the Count Syntax



ANALYZING SALES DATA AT CHIPOTLE

- Sales Data (Chips)
- Clean up: Chipotle's menu lists six different types of chips, however, our data file contains eight types of chips.
  - "Chips" and "Side of Chips" can be combined into "Chips"
  - "Chips and Fresh Tomato Salsa" and "Chips and Mild Fresh Tomato Salsa" can be combined into "Chips and Fresh Tomato Salsa"

Types of Chips	Frequency
Chips	=COUNTIFS($C$2:$C$10000,F5)
Chips and Fresh Tomato Salsa	=COUNTIFS($C$2:$C$10000,F6)
Chips and Guacamole	=COUNTIFS($C$2:$C$10000,F7)
Chips and Mild Fresh Tomato Salsa	=COUNTIFS($C$2:$C$10000,F8)
Chips and Roasted Chili Corn Salsa	=COUNTIFS($C$2:$C$10000,F9)
Chips and Tomatillo-Green Chili Salsa	=COUNTIFS($C$2:$C$10000,F10)
Chips and Tomatillo-Red Chili Salsa	=COUNTIFS($C$2:$C$10000,F11)
Side of Chips	=COUNTIFS($C$2:$C$10000,F12)

What do we notice about the naming conventions of each type of chip?

Types of Chips	Frequency
Chips	=G5+G12
Chips and Fresh Tomato Salsa	=G6+G8
Chips and Guacamole	=G7
Chips and Roasted Chili Corn Salsa	=G9
Chips and Tomatillo-Green Chili Salsa	=G10
Chips and Tomatillo-Red Chili Salsa	=G11

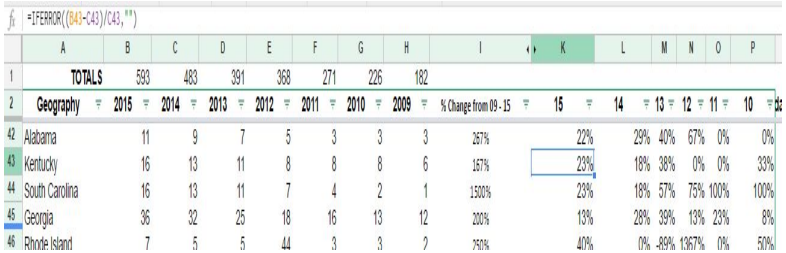
Types of Chips	Frequency
Chips	312
Chips and Fresh Tomato Salsa	111
Chips and Guacamole	479
Chips and Roasted Chili Corn Salsa	40
Chips and Tomatillo-Green Chili Salsa	74
Chips and Tomatillo-Red Chili Salsa	68

ANALYZING SALES DATA AT CHIPOTLE

- **Sales Data (Chips)**
- Now that we have a frequency distribution, let's visualize it. A **bar graph** displays the frequency of variables, represented by a series of rectangles of equal widths.
  - When creating a bar graph, it is useful to include a title, label the axes, and to use color to distinguish important features of the graph (one color will do in this case).
  - Create the bar graph by highlighting the table in Google Sheets and selecting the chart icon.
  - What percentage of orders contain "chips and guacamole"?
    - Referencing the "Sales Data (Sorted)" sheet find the max order #
      - =max(Sales Data (Sorted)A2:A1000) = 1834
      - =(479/1834) x 100= 26.12%

ANALYZING - FIND STATES W/ HIGH RECENT GROWTH

- **2009 - 15 Properties**
- We decide we want to look at opening up a restaurant in a state that has seen high growth over the past 2 years
- Make a copy of 2009 - 15 Properties
  - Highlight columns j:N and insert columns to left (right click)
- Input “15” in cell K2 14 in cell L2 etc.
  - Find year over year growth rates CELL K2 (=IFERROR(B2-C2)/C2,””)
    - Drag across to capture all
- Calculate Average 2 year growth rate
  - Only want it for states that had restaurants in 2013
    - =IFERROR(IF(D31>0,AVERAGE(K31,L31),""),"")
  - Sort Largest to smallest and filter out blanks and 0%



	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P
1	TOTALS	593	483	391	368	271	226	182							
2	Geography	2015	2014	2013	2012	2011	2010	2009	% Change from 09 - 15	15	14	13	12	11	10
42	Alabama	11	9	7	5	3	3	3	267%	22%	29%	40%	67%	0%	0%
43	Kentucky	16	13	11	8	8	8	6	167%	23%	18%	38%	0%	0%	33%
44	South Carolina	16	13	11	7	4	2	1	1500%	23%	18%	57%	75%	100%	100%
45	Georgia	36	32	25	18	16	13	12	200%	13%	28%	39%	13%	23%	8%
46	Rhode Island	7	5	5	4	3	3	2	250%	40%	0%	-89%	1367%	0%	50%

TOTALS		
Geography	% Change from 09 - 15	2 YEAR GROWTH
Louisiana		75.0%
Arkansas		62.5%
Utah	100%	50.0%
Idaho		41.7%
Maine		41.7%
France		41.7%
Nevada	118%	36.2%
Washington	150%	32.9%
North Carolina	255%	30.5%
Pennsylvania	259%	28.9%
Canada	1000%	28.6%
Alabama	267%	25.4%
Kentucky	167%	20.6%

SUMMARIZING FINDINGS

ACTIVITY

The GA Analysts have been invited to Chipotle's Headquarters in Denver, CO to share the results of their analysis. While preparing for your presentation to senior managers, you bump into Steve Ells, Chipotle's founder in the hallway. He's excited and asks, "I'm on my way to another meeting, but what did you find?"

Working in pairs, and using our format for a recommendation based on data, what *one* recommendation do you share with him? Remember, you'll only have his attention for a minute, at most!

1. Summarize (stating the results)
2. Recommend (interpret the results and draw a conclusion)
    a. What is this data saying?
    b. What can we infer from this data?
3. Future Analysis (consider the shortcomings of the data and plan for future analysis)

CONCLUSION

CONCLUSION

- Today, we've seen the Data Analytics workflow in action!

- We've identified questions to answer with data; we took the necessarily steps (obtaining, understanding, and preparing our data) prior to starting our analysis; then, we performed analysis. Finally, we've used our analysis to form actionable insights from our data.

CONCLUSION

- Want more help?
  - Google Sheets Help Documentation by Google

- Want to continue your data journey at GA?

  - SQL Bootcamp
  - Data Analysis Circuit (Part-Time Online)
  - Data Analytics (Part-Time or 1-Week Accelerated)
  - Data Science (Part-Time or Immersive)

CITATIONS

- Chipotle Mexican Grill, Inc., Annual Reports, 2009 - 2015: http://ir.chipotle.com/phoenix.zhtml?c=194775&p=irol-reportsAnnual
- New York Times, “At Chipotle, How Many Calories Do People Really Eat?”: http://www.nytimes.com/interactive/2015/02/17/upshot/what-do-people-actually-order-at-chipotle.html

ASK ME ANYTHING

Please don’t forget the survey!

THANK YOU