WELCOME TO DATA ANALYTICS

Learning Objectives

- GA Orientation
- Overview of Course
- Value of Data and Business
- Analytics Workflow
- Review of prework

Value of Data

DAY 1 ORIENTATION

- **AGENDA**
  - ➔ Welcome + Introductions
  - ➔ Student Experience
  - ➔ Course Expectations
  - ➔ Course Tools
  - ➔ Start Learning!

HELLO.

- # INSTRUCTOR NAME - INSTRUCTIONAL LEAD
  - ➔ Position
  - ➔ Advanced Analytics Analyst
  - ➔ Company
  - ➔ Costco Wholesale
  - ➔ @matthewmorris (Slack)
  - ➔ Matthewmorris.Da@gmail.com

HELLO.

- **MARIA SENGLE**
  - ➔ Operations Manager
  - ➔ Campus / Course Logistics

  - ➔ @maria.sengle (Slack)
  - ➔ seattle_production@ga.co

HELLO.

- **STEPHANIE CARUSO**
  - ➔ Education Programs
  - ➔ Marketing & Workshops

  - ➔ @scaruso (Slack)
  - ➔ stephanie@generalassemb.ly

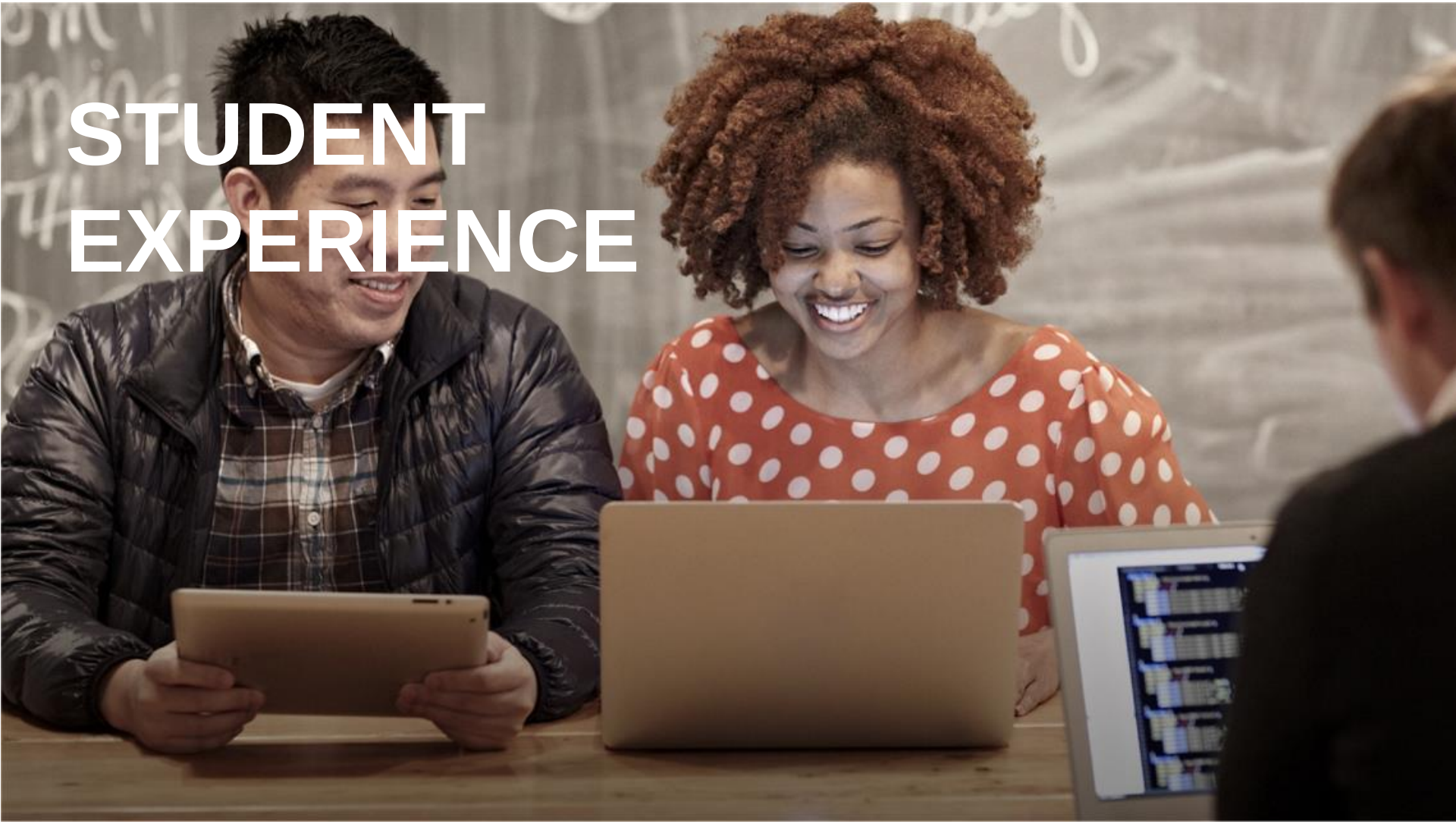
HELLO.

- **BRYNA LIEBERMAN**
  - ➔ Education Programs Lead
  - ➔ Instructor Support
  - ➔ @bryna (Slack)

WHO ARE YOU?

- **INTRODUCTIONS**
  - ➔ What's your name?
  - ➔ What are you up to these days?
  - ➔ Why are you taking this course?

  - ➔ Fun Fact!
    - ◆ *Something you <u>don't</u> normally share i.e. guilty pleasure i.e. tell us your secrets*


STUDENT
EXPERIENCE

SEATTLE CAMPUS

- **HOUSEKEEPING**
  - ➔ Student Handbook: campus guidelines, hours, etc. This is your home for the next 10 weeks!
  - ➔ Shared space: talk to your neighbors, clean up, make friends
  - ➔ Kitchen (coffee, tea, snacks, please label your food!)
  - ➔ Bathrooms: keys at front desk

STUDENT SUPPORT

FRONTLINES: CAMPUS FACILITIES

Door access, printing needs, heat/AC, chargers, etc.

MARIA: COURSE COUNSELOR / CHEERLEADER

Here to help with any course logistics/questions

(payments, tools/systems, etc.)

INSTRUCTOR: SUBJECT MATTER EXPERT

Yoda. Miyagi. Gandolf.

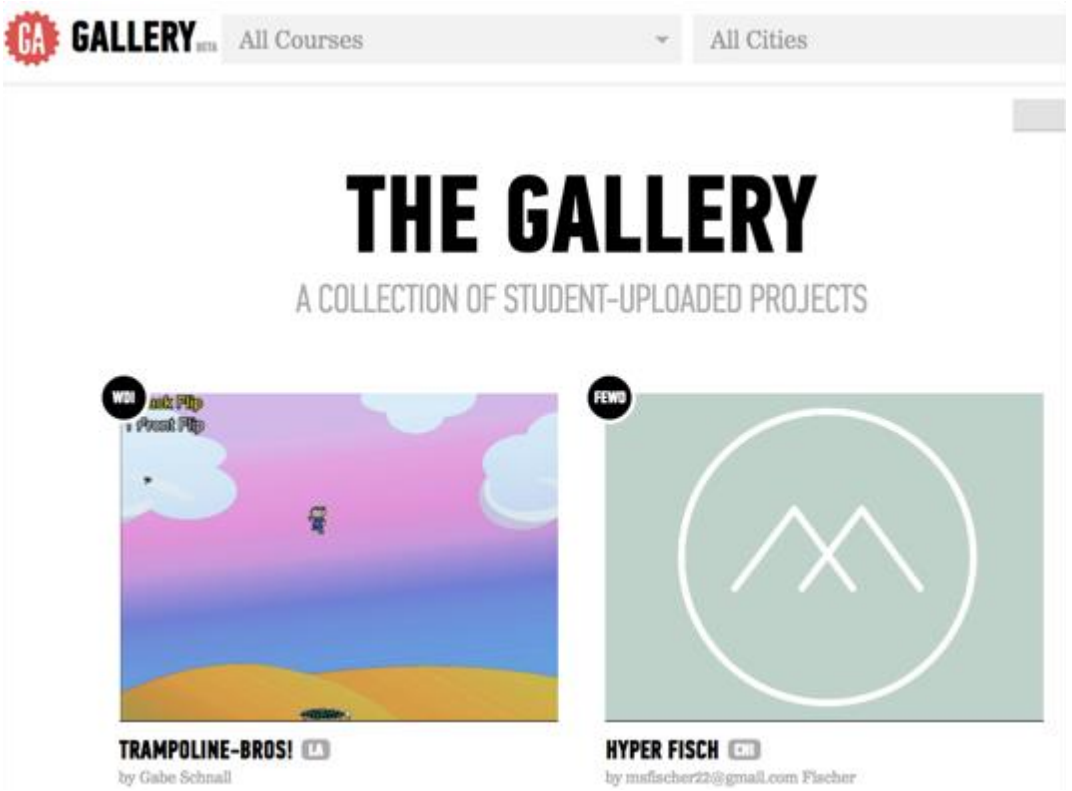
STUDENT SUPPORT

- **FRONT LINES**
  - ➔ 206.258.7033
  - ➔ Slack: @frontlines
  - ➔ seafrontdesk@ga.co

BE INSPIRED!

- **GA GALLERY**
  - ➔ The GA Gallery is our global showcase of student projects: all courses, all campuses
  - ➔ Get inspiration
  - ➔ Post YOUR project!

- gallery.ga.co




COURSE
EXPECTATIONS

COURSE EXPECTATIONS

- **ADULT LEARNING**
  - ➔ Be on time (i.e. early)
  - ➔ Complete your assignments and submit them on time
  - ➔ Participate + ask questions
  - ➔ Share with your peers
  - ➔ Make friends :)

  - ➔ You will **get out** what you **put in** to the class

COURSE EXPECTATIONS

- **ATTENDANCE**
  - ➔ Let your instructor know if you will be absent (hint: DM on Slack!
  - ➔ Make plans to catch up if you know you are going to miss class
  - ➔ 15 minutes late = 1 Tardy
  - ➔ 3 Tardies = Absence *(WA State Law)*

IF YOU HAVE TO MISS A CLASS:

CLASS RESOURCES		CLASSMATES		INSTRUCTOR
➔ Look over the slides ➔ Attempt the homework		➔ Grab coffee with your peers ➔ Borrow notes from class		➔ Attend office hours after reviewing materials and be ready with specific questions

LETTER OF COMPLETION REQUIREMENTS

ATTENDANCE: MISS NO MORE THAN 3 LESSONS

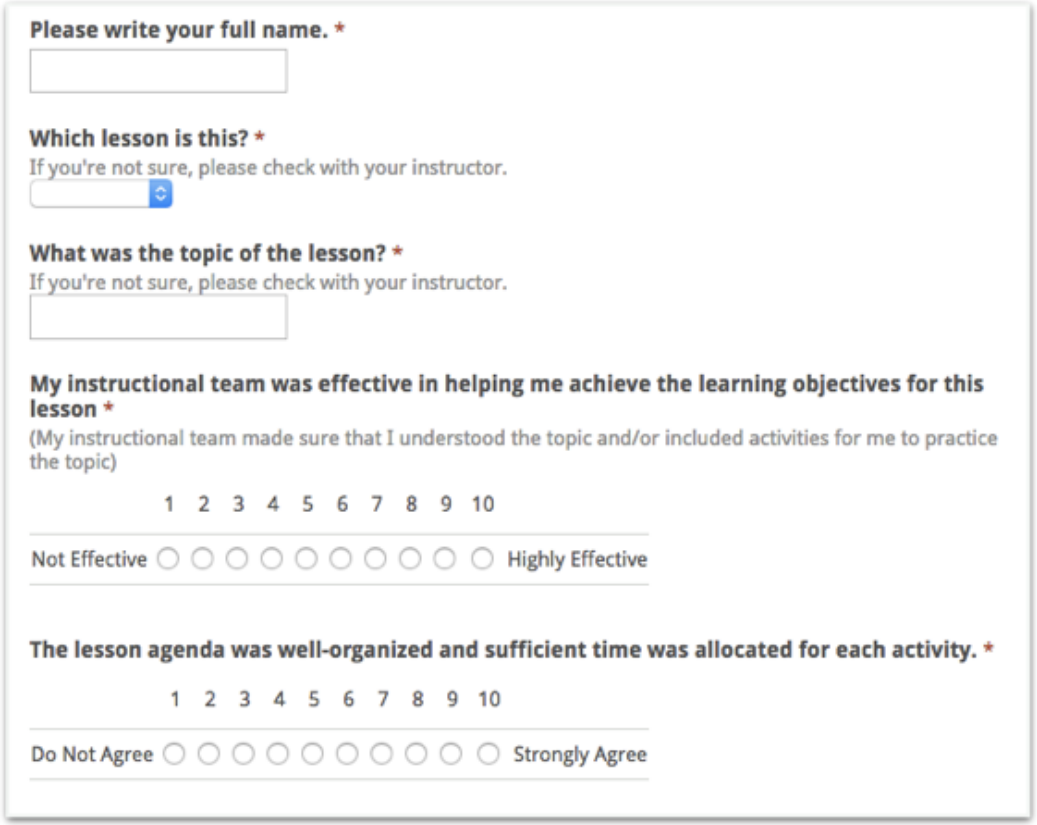
ASSIGNMENTS: COMPLETE >80% OF HOMEWORK +

MEET CRITERIA FOR, PRESENT & SUBMIT FINAL PROJECT

FEEDBACK: PARTICIPATE IN MID- & END-OF-COURSE FEEBACK SURVEYS

HOW'S IT GOING

- **FEEDBACK**
  - ➔ Daily Exit Tickets (feedback + reflection)
  - ➔ Mid-Course Survey
  - ➔ End-of-Course Survey



Please write your full name. *

Which lesson is this? *
If you're not sure, please check with your instructor.

What was the topic of the lesson? *
If you're not sure, please check with your instructor.

My instructional team was effective in helping me achieve the learning objectives for this lesson *
(My instructional team made sure that I understood the topic and/or included activities for me to practice the topic)

1 2 3 4 5 6 7 8 9 10

Not Effective ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Highly Effective

The lesson agenda was well-organized and sufficient time was allocated for each activity. *

1 2 3 4 5 6 7 8 9 10

Do Not Agree ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Strongly Agree

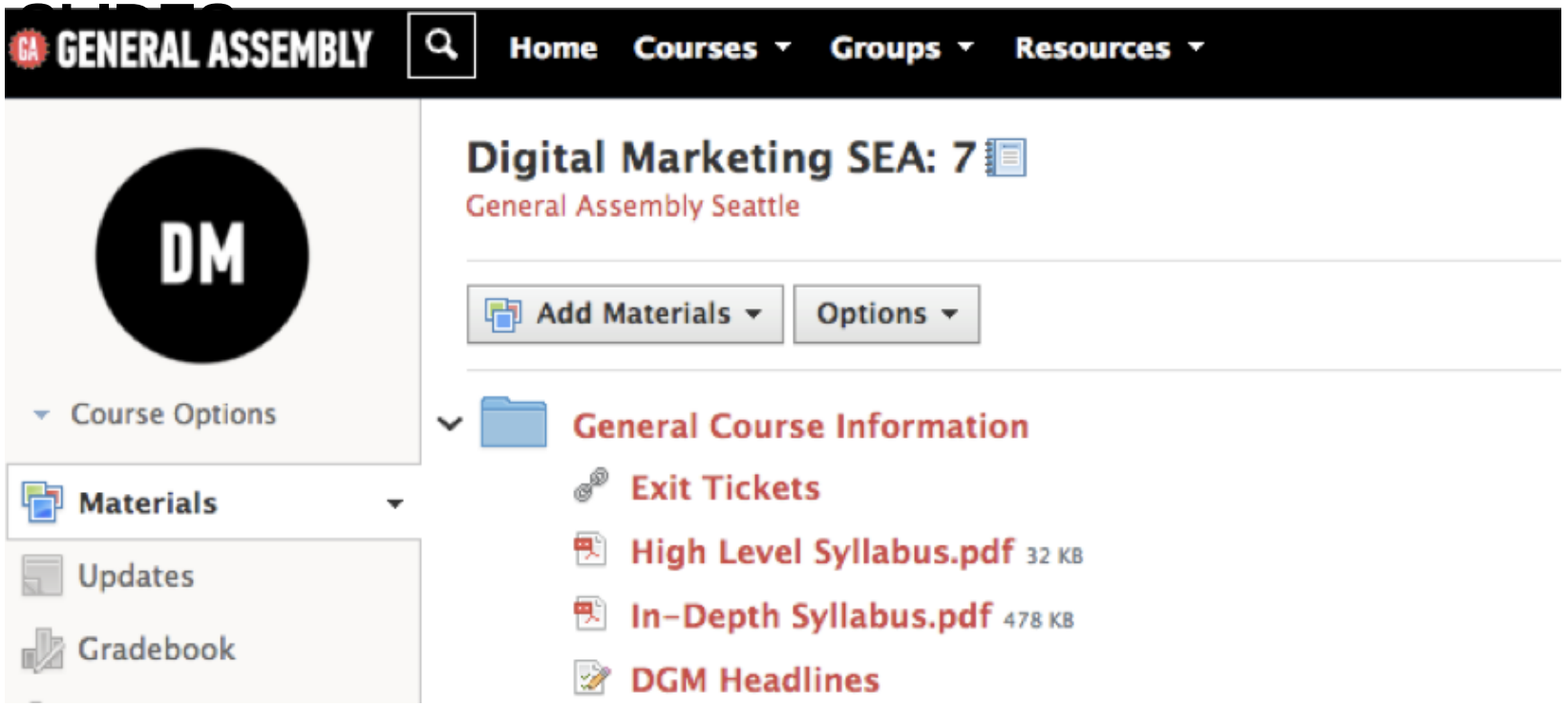

TOOLS
FOR SUCCESS

SLACK

- **INSTANT MESSENGER**
  - ➔ Class communication + GA announcements
  - ➔ Part-time student community: #general, #frontlines, #random, #happyhour
  - ➔ Direct Messages
  - ➔ Desktop / mobile app!

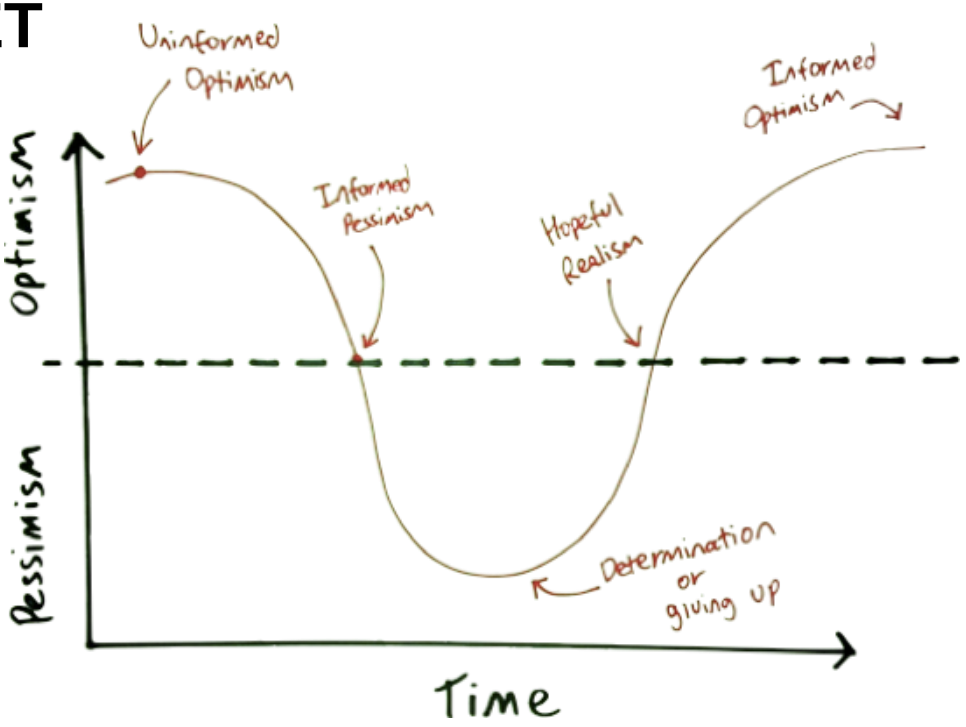
SCHOOLOGY (SKOOL-UH-GEE)

- **ASSIGNMENTS,**



TIPS FOR SUCCESS

- **THE LEARNING ROLLERCOASTER + GROWTH MINDSET**



GENERAL ASSEMBLY

FUNDAMENTALS OF DATA AND EXCEL

Matthew Morris

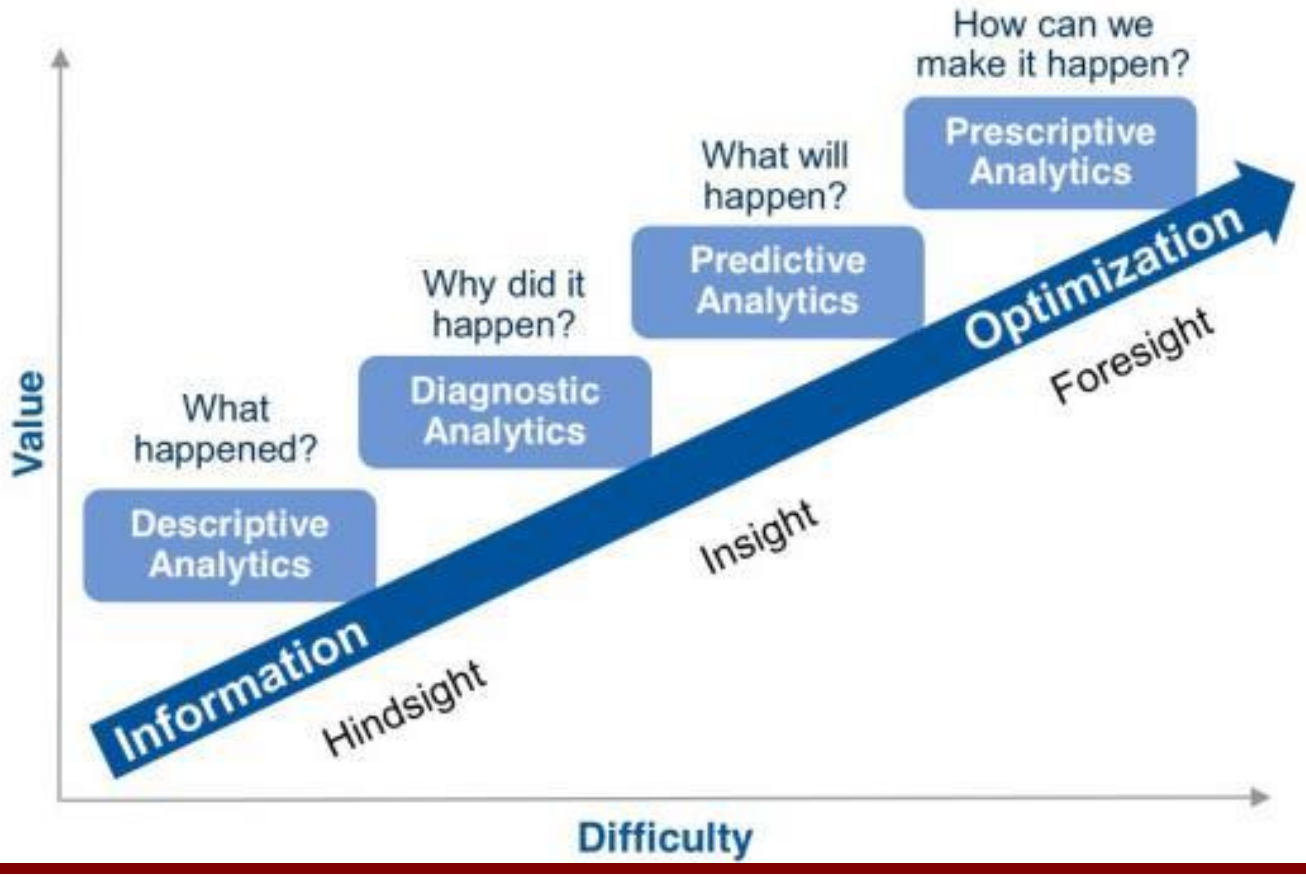
Git: Morrisdata

MatthewMorris.DA@gmail.com

INTRODUCTION TO THE COURSE

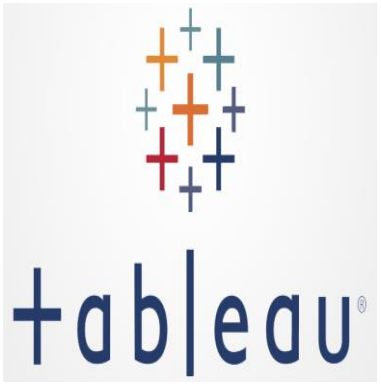
- Now let's meet you!
  - Where do you study or work?
  - What challenges have you faced when "tackling" a dataset or producing a report based on findings?
  - What software have you used when working with data?

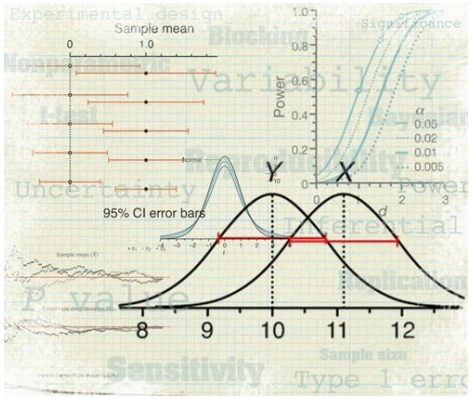
DATA ANALYTICS/DATASCIENCE



Overview of Course

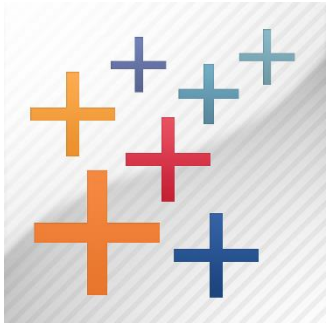
Course Objectives





Value of Data

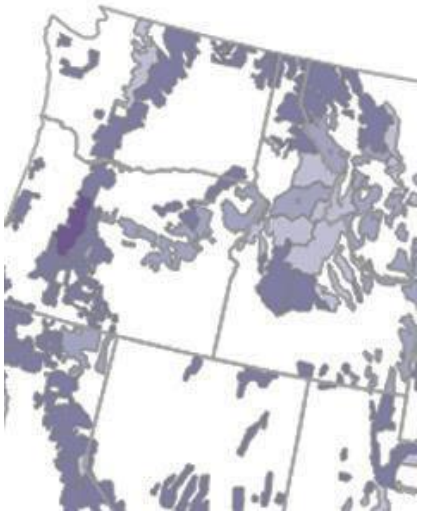
Overview of Course



amazon web services

Value of Data

Course Content



Value of Data

Course Content

WEEK 1	2/21/2017	Fundamentals of Data and Excel
	2/23/2017	Data Referencing in Excel
WEEK 2	2/28/2017	Data Aggregation in Excel
	3/2/2017	Data Narratives
WEEK 3	3/7/2017	AirBnB Investment Presentations
	3/9/2017	Fundamentals of Databases and SQL
WEEK 4	3/14/2017	Filtering and Aggregating in SQL
	3/16/2017	Querying Large Databases
WEEK 5	3/21/2017	Creating Multiple Join Relationships
	3/23/2017	Data Aggregation in SQL
WEEK 6	3/28/2017	Using Subselects in SQL
	3/30/2017	Applying Functions in SQL
WEEK 7	4/4/2017	Creating Local Database
	4/6/2017	Mozilla Firefox Presentations
WEEK 8	4/11/2017	Fundamentals of Tableau
	4/13/2017	Visualization and Text Manipulation
WEEK 9	4/18/2017	Calculations and Analysis in Tableau
	4/20/2017	Dashboarding in Tableau
WEEK 10	4/25/2017	Flex day, review, catch up, workshop
	4/27/2017	Presentation

INTRODUCTION TO THE LESSON

PRE-WORK

- Congratulations on conquering the onboarding task!
  - What was your favorite topic that you covered?
  - What is one topic that you had trouble with?

Cleaning Data

Normalize

Remove duplicates

Hide unwanted columns

Recalculate data to add value

Cleaning Data

Case text inconsistencies
Spaces
Non-print characters
Numbers and number signs
Dates, times, and custom formats

Demo Business Scenario

- We will be working for the State of Washington's Governor's Office as a policy analyst.
- Policy analysts often use many different data sources in order to evaluate policy decisions and make recommendations on how to allocate resources, hold entities accountable, and more.
- As policy analysts, we will be using data from the American Community Survey (ACS), which is a random survey given each year to residents of the United States.

Demo Business Scenario

- It is a sample, which means that not everyone is required to respond (unlike the census, which occurs every 10 years).
  - During a census, everyone is asked to respond.
  - Because of this burden, the census happens once a decade.
  - Sampling is used in the off-years to provide *estimated* data about the population.
  - For the ACS, approximately 1/36 households are asked to respond.
  - Because it is a sample, each variable is an estimate that has a degree of error associated with it based on number of respondents, sampling strategy, and more.

Demo Business Scenario

- Here is a diagram from the United States Census Bureau on how the ACS works:

http://www.census.gov/programs-surveys/acs/about/how-the-acs-works.html

INDEPENDENT PRACTICE: CLEANING OUR DATASET

Practice: Cleaning your Data Set

- Here is important information you need to know about our dataset:
  - Many of the ACS tables have data aggregated by census tract. Census tracts are small areas, sometimes as small as a few blocks in a densely populated area such as Manhattan, that the ACS uses for tabulation. Each census tract has an ID, and that is the “id” field in our dataset.
  - As we are working for the State of Washington, our dataset only includes census tracts in Washington.

Practice: Cleaning your Data Set

- Here is important information you need to know about our dataset:
  - Sometimes data is reported as a total of those counted. For example, the dataset has Total Population of census tract and number of females, but it does not have the percentage of people who are female.
  - Other times, data is reported as a percentage. For example, the unemployment rate is provided in the dataset.

In order to be able to perform analysis using the ACS dataset, we will need to make some changes and do some exploration.

Practice: Cleaning your Data Set

EXERCISE

DIRECTIONS

1. Open 2014_acs_select_WA.xlsx
2. Based on your experience, choose either the BASE or STRETCH tab to complete (60 min).

You may work with a partner, checking in with each other after answering each question.

DELIVERABLE

Complete BASE or STRETCH tab in 2014_acs_select_WA.xlsx

GUIDED PRACTICE: REVIEWING OUR SOLUTIONS

Review Solutions

- Create percentage columns where possible.
  - To do this, for each estimate column that is not already a percentage or a rate, we should pair them up in terms of numerator and denominator.
  - For example, column E and F are Total Population and population of Males, respectively.
    - Create a new column with header “% Male of Population”.
    - Enter the formula “=F2/E2” in the first cell under the header (row 2).
    - Double-click the square in the bottom-right of the cell to copy this formula down all of the rows.

Review Solutions

- Convert all percentage columns to values between 0-1 (inclusive) with a format of '00.00%'.
  - There are two options:
    - Create a new column which is equal to the old column but divided by 100.
    - Use Paste Special's "Divide" feature.

Review Solutions

- Remove all rows with no data.
  - What to do with rows that are completely empty? This is always decided on a case-by-case basis, as empty data is sometimes very useful. But for our purposes, when we are making policy recommendations about people who live in these census tracts, we can delete these rows.
  - Always document what you delete and make a note of your rationale.

Review Solutions

- Create a common code for cells with no data.
  - It is also nice/necessary to have empty or null values coded in a consistent way. This dataset is already pretty consistent with both blank and '-' cells. However, one quirk of Excel is that, if there is textual data amongst numeric data, it can really mess up how that data is plotted. For our purposes, let's recode all empty cells to be blank.

Review Solutions

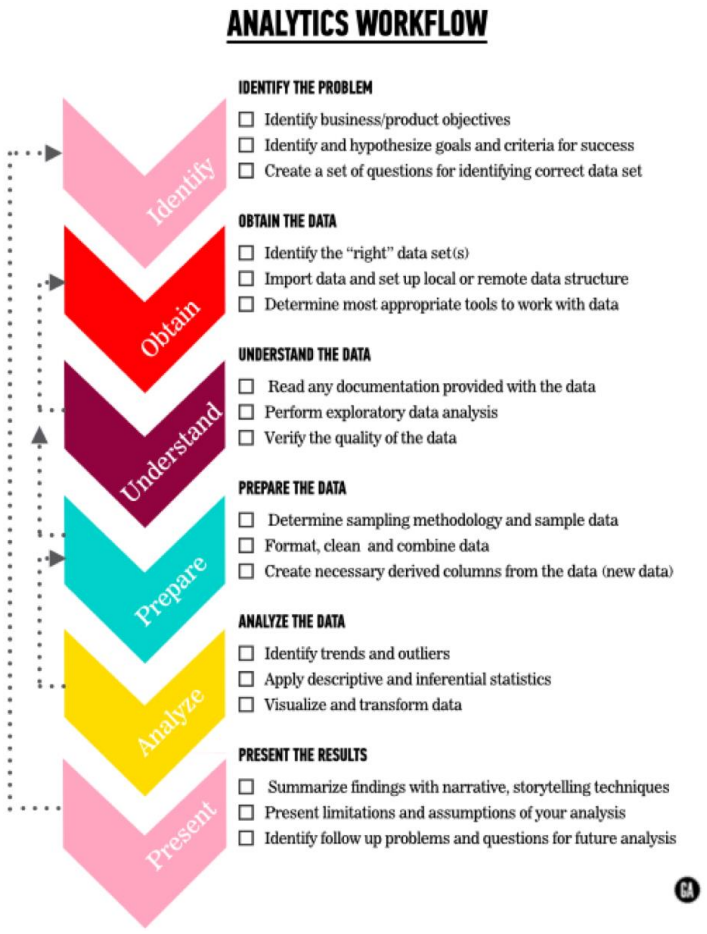
- Is there any data that could be erroneous? If so, what are our options (if any)?
  - Click through each column's filter drop-down and take a look at the values. Do you see anything that doesn't pass a gut check?
  - For now, there is not much we can do except document our findings. Make a note about the age field and others that might be questionable.
    - To help us determine if the median age is unreasonable, it can be helpful to look up information about the questionable tract to see if it actually makes sense.

Review Solutions

- What were some interesting findings?
  - Exploratory data analysis is always helpful for finding potentially erroneous data, as well as obtaining an understanding of what data you have in your dataset.
  - Scatterplots are a great tool for looking at relationships between two variables.

Review Solutions

- A significant amount (usually more than half, if not more than 75%) of any project involves cleaning data and basic exploratory analysis to become familiar with the data. Only after you understand the nuances of the dataset and have structured it appropriately can you move on to the more advanced steps of learning from the dataset.



Project 1

Prompt: You are doing work for a client that wishes to invest in an AirBnB hotel in Amsterdam. Before they decide to invest, they would like clear data about the AirBnB performance in that specific market, what property types receive the most positive reviews, which neighborhoods host the most listings, how much revenue successful hosts generate, and so forth…

Value of Data

Project 1

Five-minute Presentation - during Lesson 6

Business needs as per your interpretation of the scenario;
Data selected from the original file;
Cleaning methods used to remove erroneous data;
Format: Google Slides or PDF (Keynote/PPT need to be exported); presentation will be given in small groups.

Value of Data

CONCLUSION

EXIT TICKET

EXIT TICKETS

Name of class : Fundamentals of Data and Excel

Question : What is the best way to approach Data analytics?

Contact: MatthewMorris.DA@gmail.com

CREDITS

CITATIONS

- The datasets used were compiled from the American Community Survey (ACS): https://www.census.gov/programs-surveys/acs/
- The datasets were downloaded directly from the American FactFinder site: http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
- All data is the from 2014 5-Year Estimate ACS.
- Summary of the ACS Data Collection: http://www.census.gov/programs-surveys/acs/about/how-the-acs-works.html

RESOURCES

- A thorough guide to the steps of data cleansing: https://www.siop.org/tip/backissues/Jan05/PDF/423_089to096.pdf
- To find these census tracts on a map, you can use this website: https://www.huduser.gov/qct/qctmap.html
  - To search, enter the portion of the ID after “US”