WELCOME TO DATA ANALYTICS

Orientation For GA and Student FAQ
Orientation For Data Analytics
Beginning of course!

Matthew Morris
Git: Morrisdata
MatthewMorris.DA@gmail.com

DAY 1 ORIENTATION

AGENDA

- ➔ Welcome + Introductions
- ➔ Student Experience
- ➔ Course Expectations
- ➔ Course Tools!
- ➔ Start of class!!!
- ➔ Intro to Data Analytics
- ➔ Data Analytic Tools
- ➔ Group Exercises (review pre-work)
- ➔ Data Analytic Workflow

DAY 1 ORIENTATION

By the end of todays class you will be able to:

- ➔ Have tools necessary to complete the course
- ➔ Describe what a Data Analyst is General principles of Data Analytics
- ➔ Use parts of whole to add value to a data set
- ➔ Navigate excel to answer questions about data

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*

HELLO.

INSTRUCTOR NAME - INSTRUCTIONAL LEAD

- Position
- Advanced Analytics Analyst, Technical Project Lead
- Company
- Costco Wholesale
- @matthewmorris (Slack)
- Matthewmorris.Da@gmail.com


STUDENT
EXPERIENCE

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.

ANNA: 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

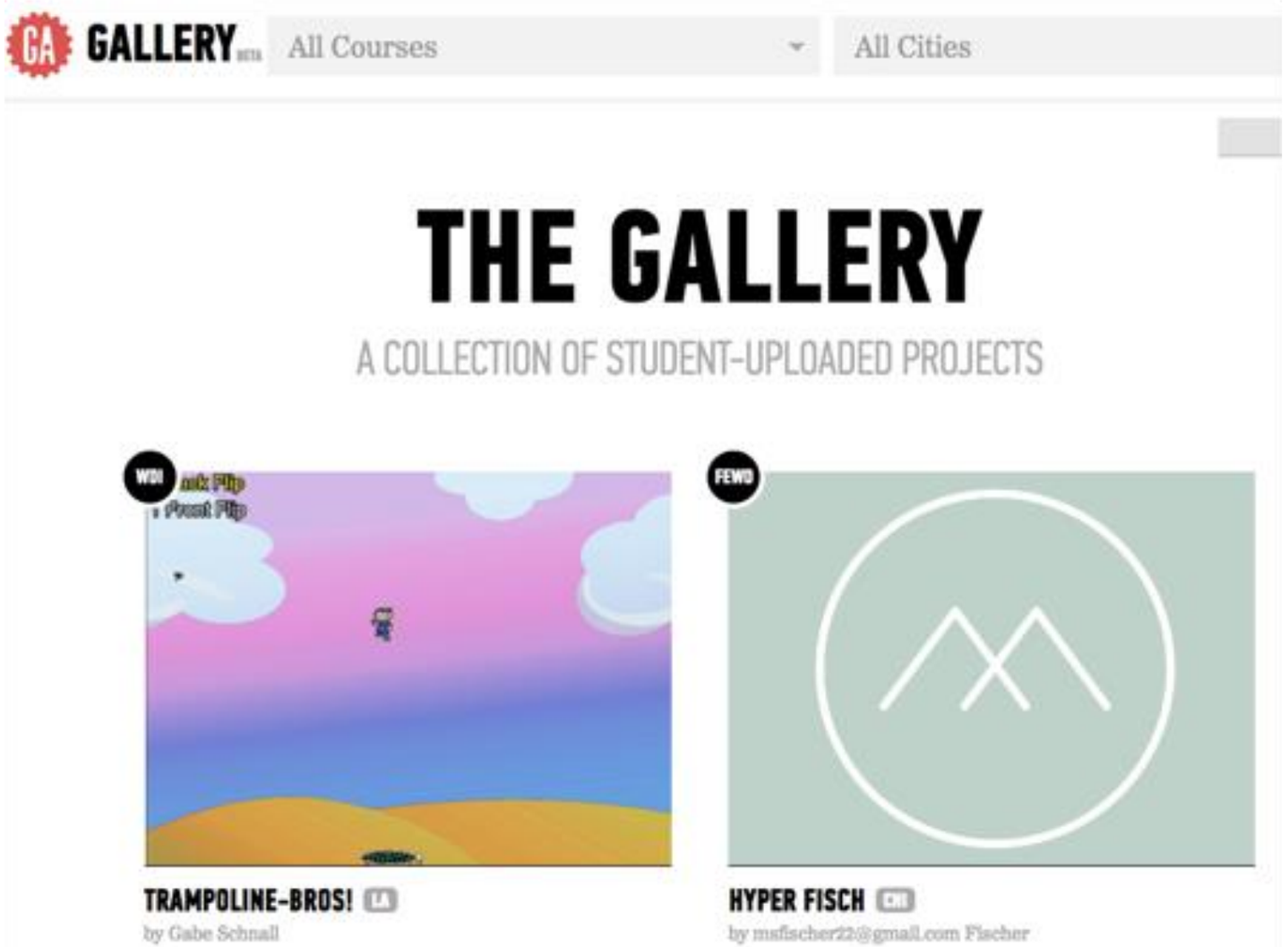
- 206.258.7033
- Slack: @frontlines
- seafrontdesk@ga.co

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

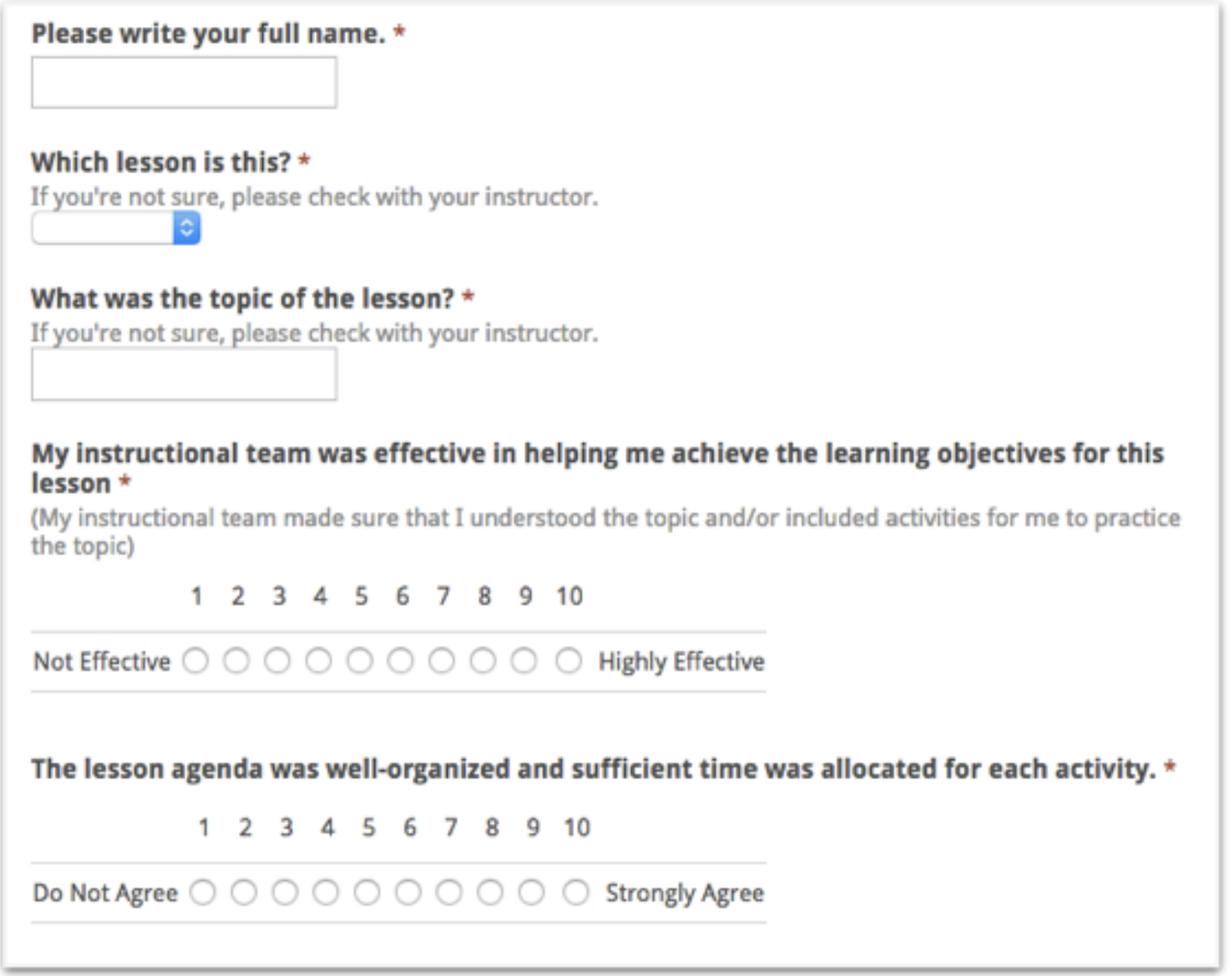
ASSIGNMENTS: COMPLETE >80% OF HOMEWORK +
MEET CRITERIA FOR, PRESENT & SUBMIT FINAL PROJECT

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!

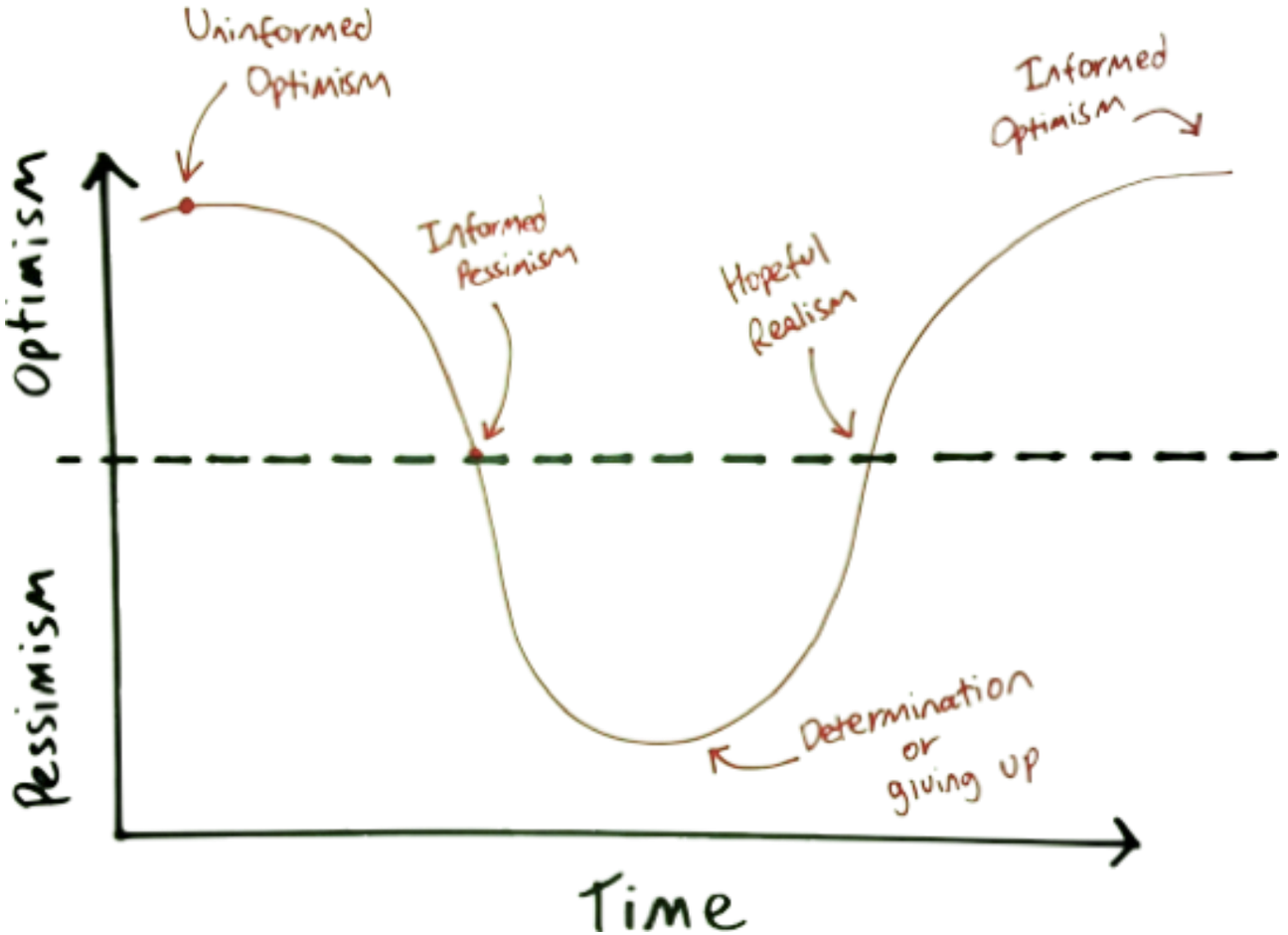
GITHUB

ASSIGNMENTS, SLIDES

github/Morrisdata/AN16

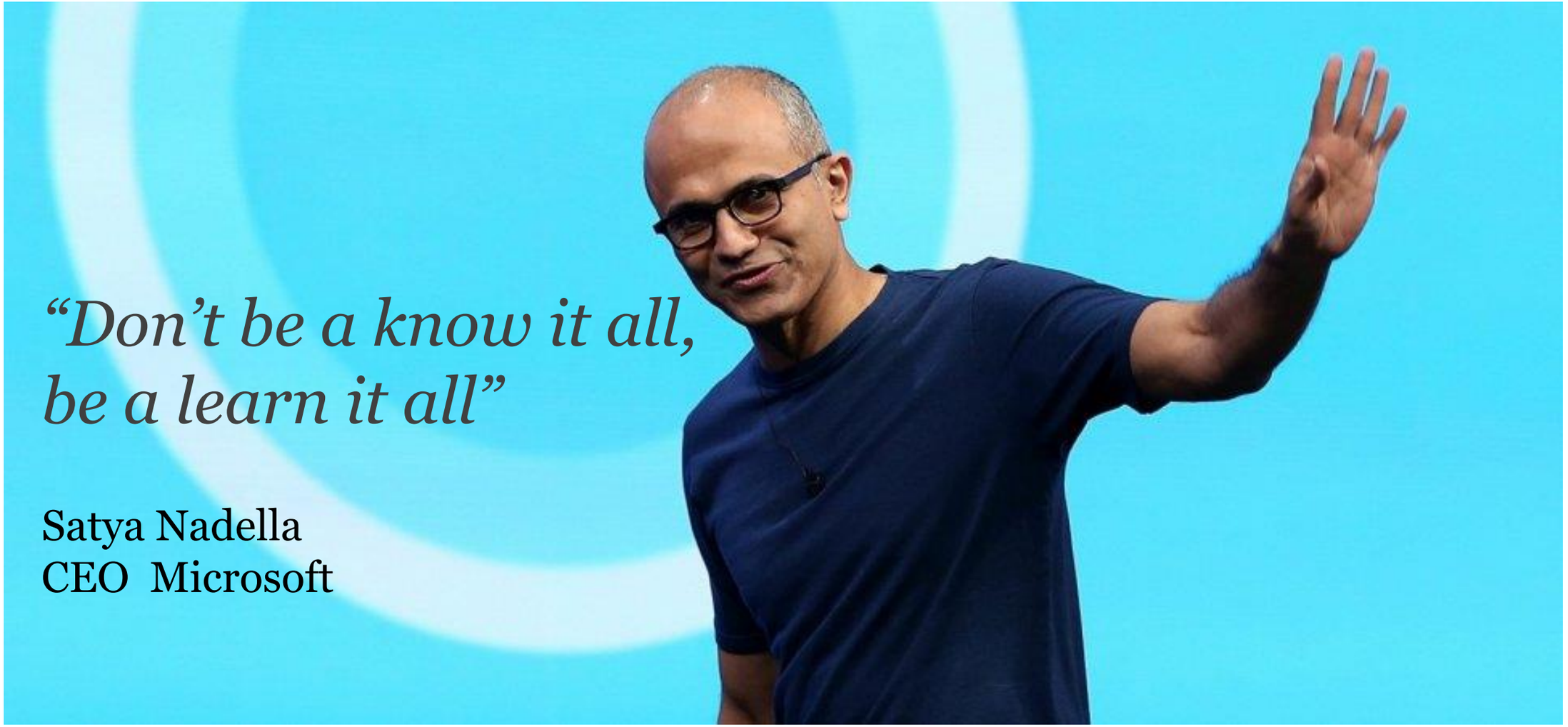
TIPS FOR SUCCESS

THE LEARNING ROLLERCOASTER + GROWTH MINDSET



TIPS FOR SUCCESS

ALWAYS BE LEARNING

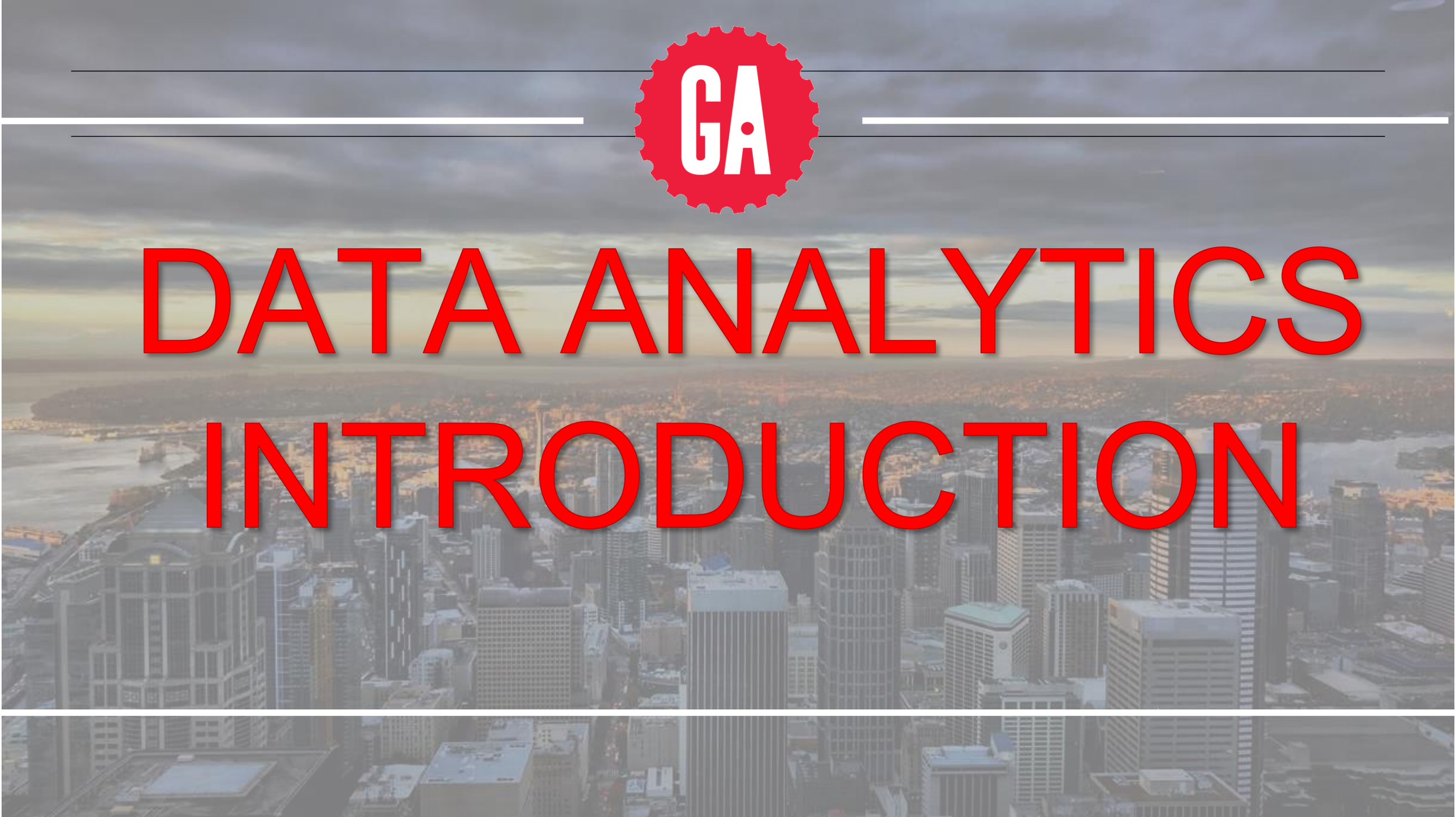


TIPS FOR SUCCESS

TIPS FOR SUCCESS

“I dedicated 1 to 3 hours a day, determined to master the material”

TIPS FOR SUCCESS


GA
DATA ANALYTICS
INTRODUCTION

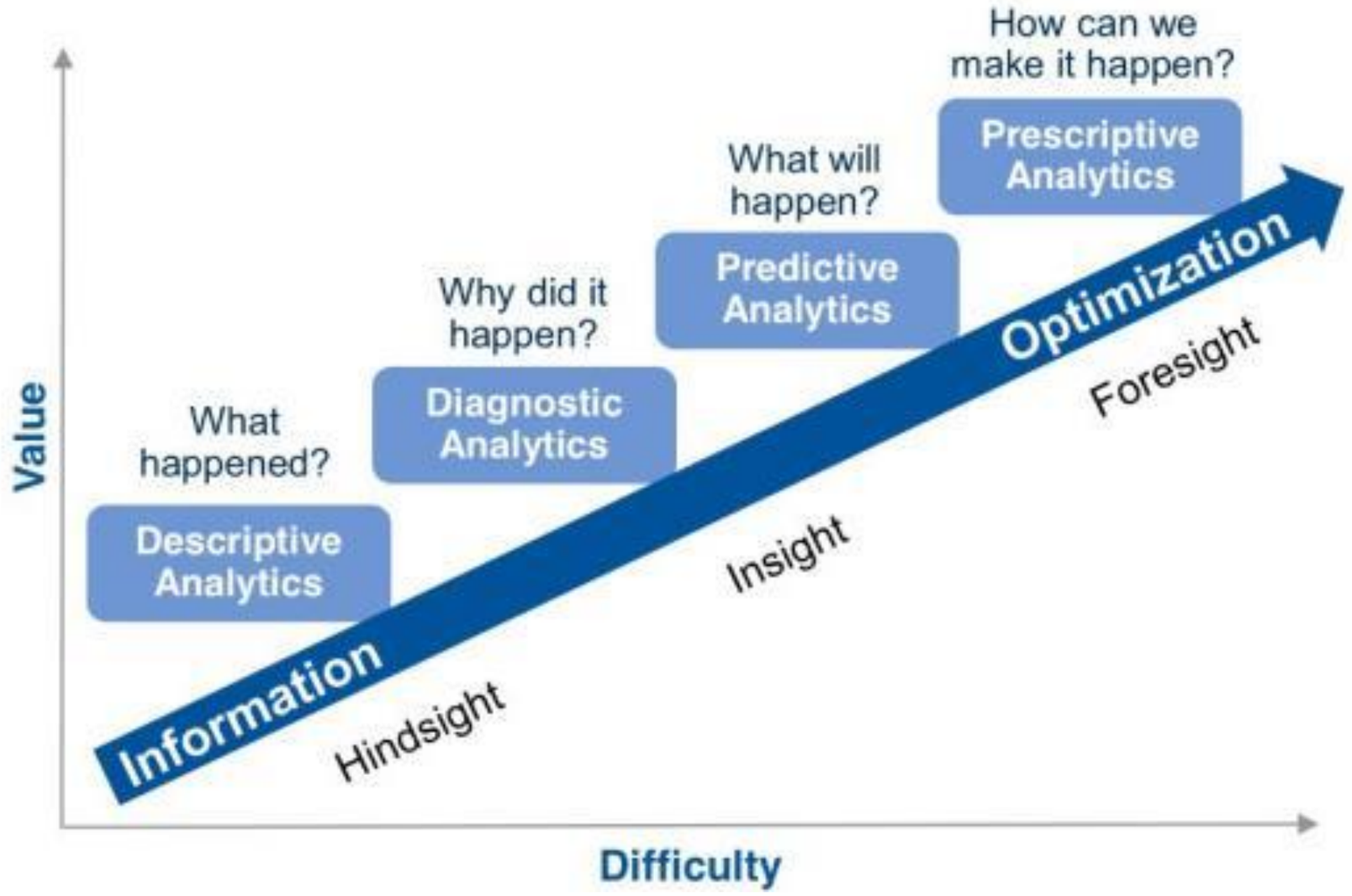
Student Objectives:

- Objective 1

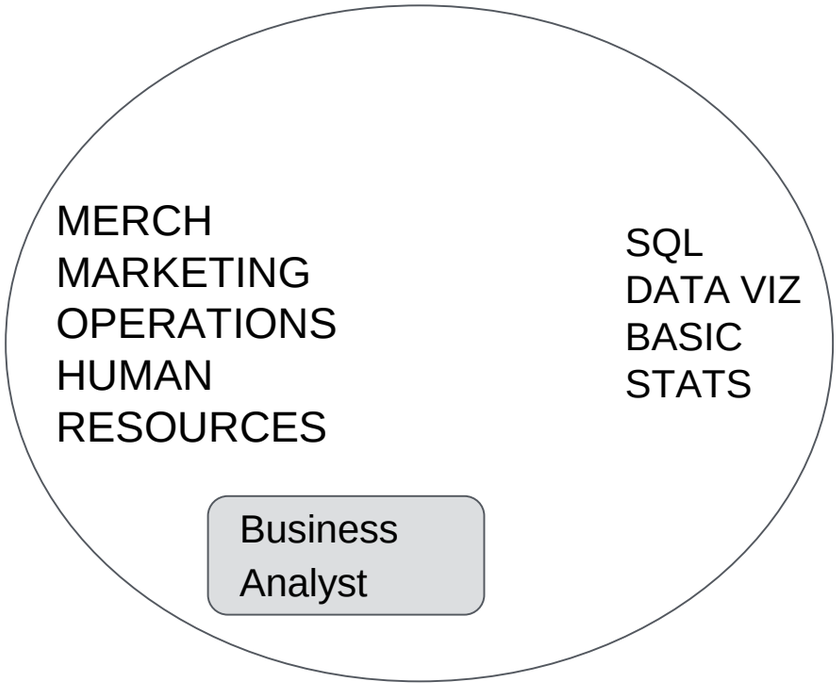
Data Analytics

- Email marketing reports
- Website analytics
- Social media analytics
- Sales trends
- Current customer data
- Sales receipts
- Media coverage of new trends
- POS records
- Inventory monitoring/tracking
- Industry surveys
- Prospective customer data
- Third-party data
- Other

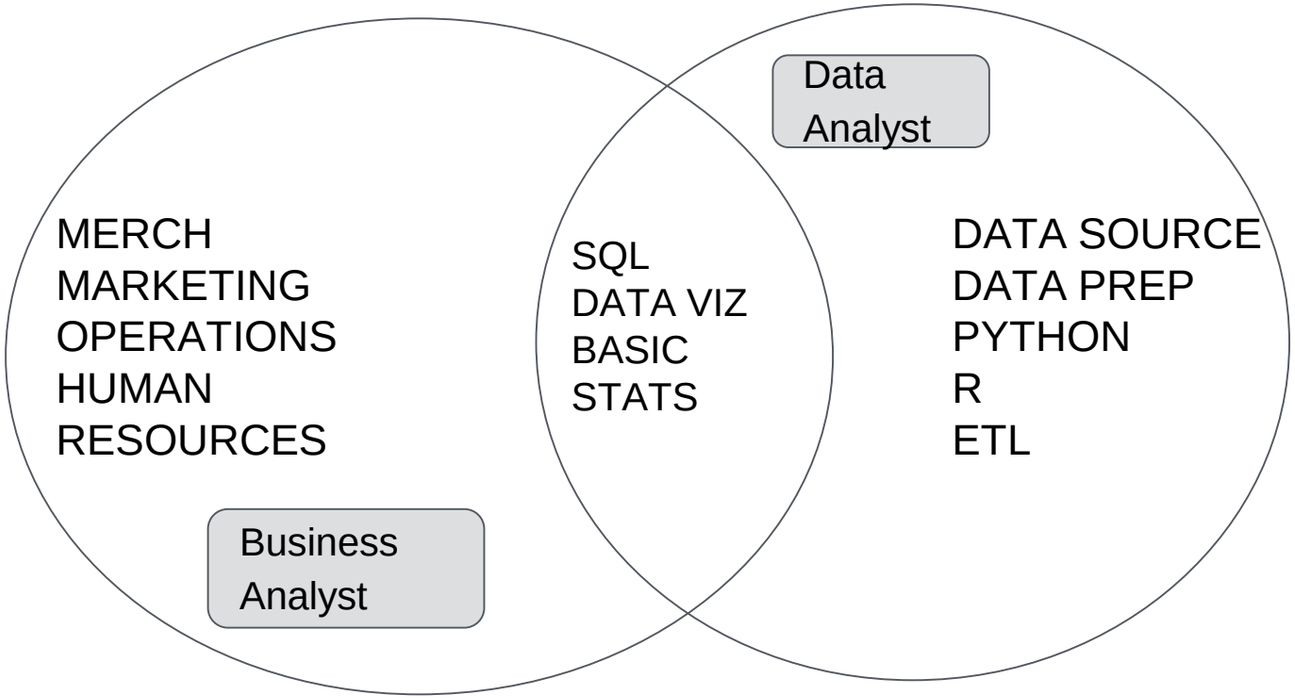
Data Analytics



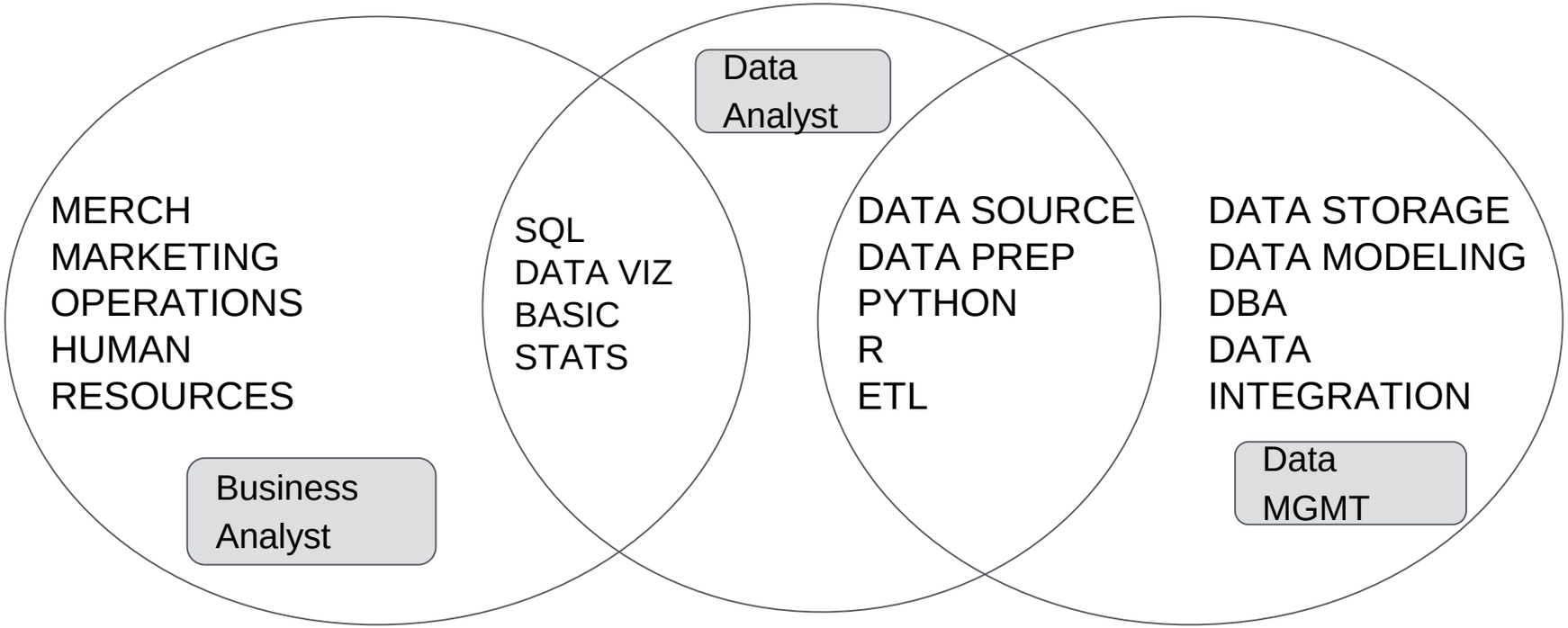
Data Analytics



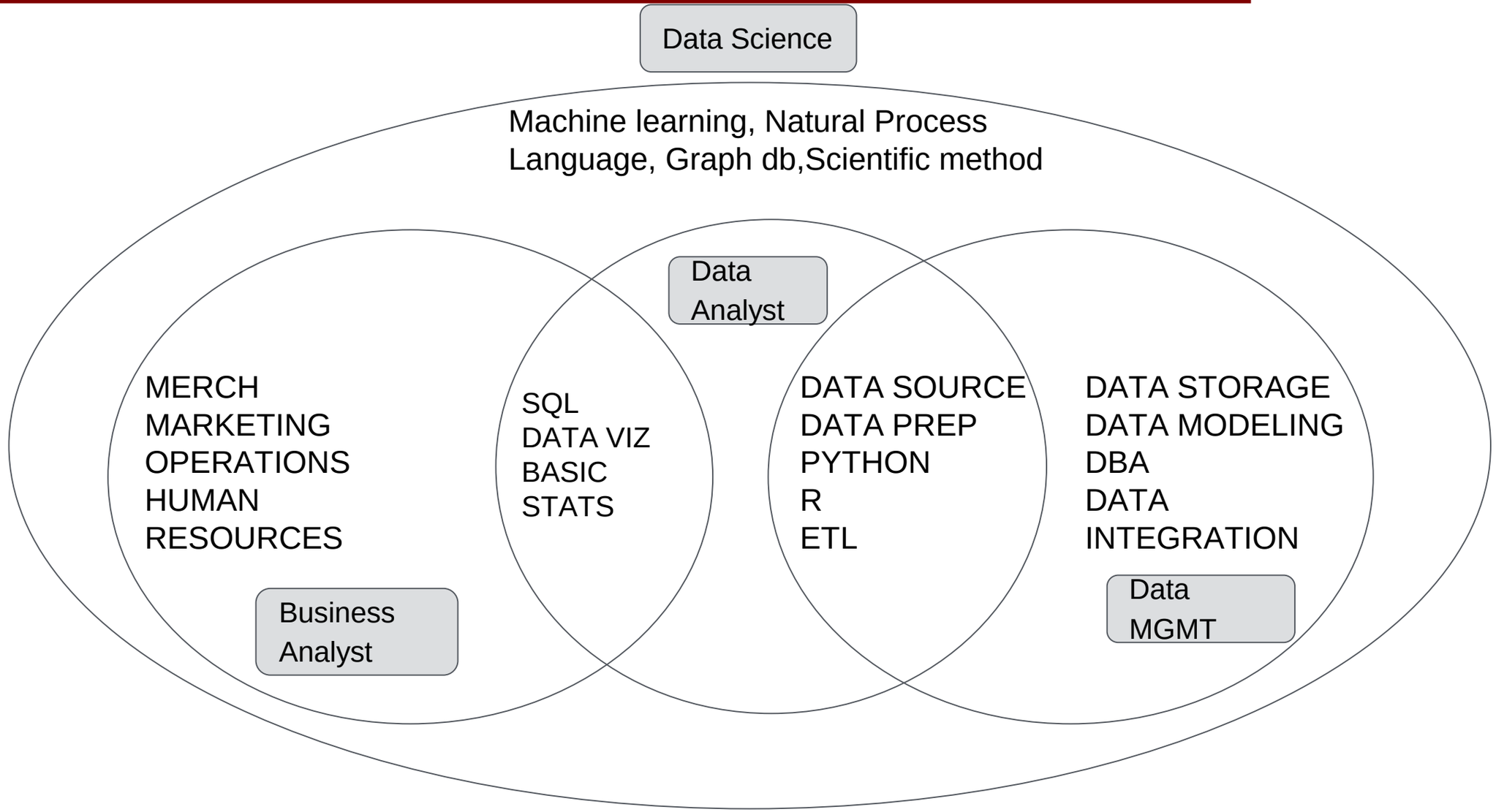
Data Analytics



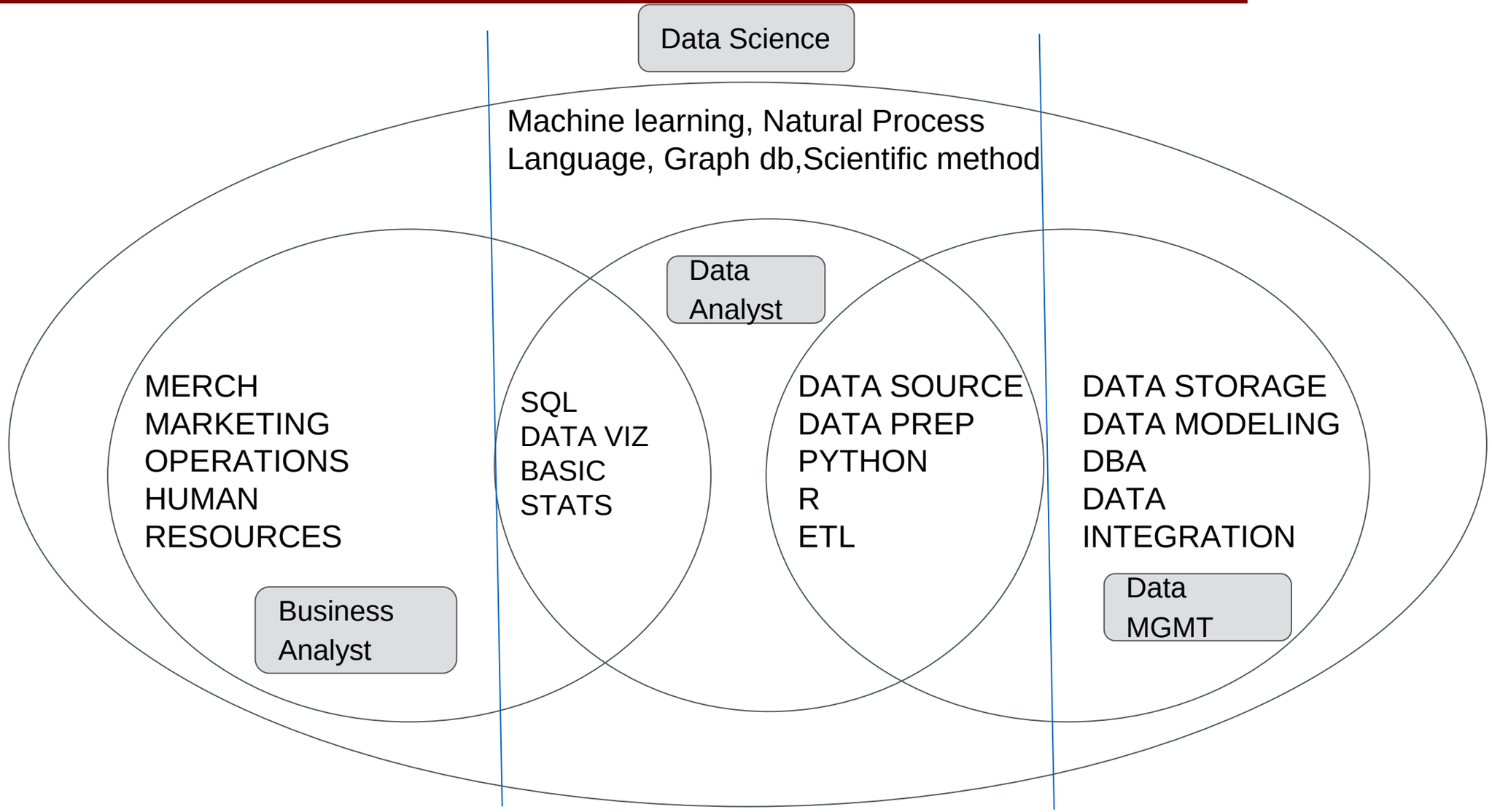
Data Analytics



Data Analytics



Data Analytics



Data Analytics

Facts and statistics collected together for reference or analysis

Things known or assumed as facts, making the basis of reasoning or calculation.

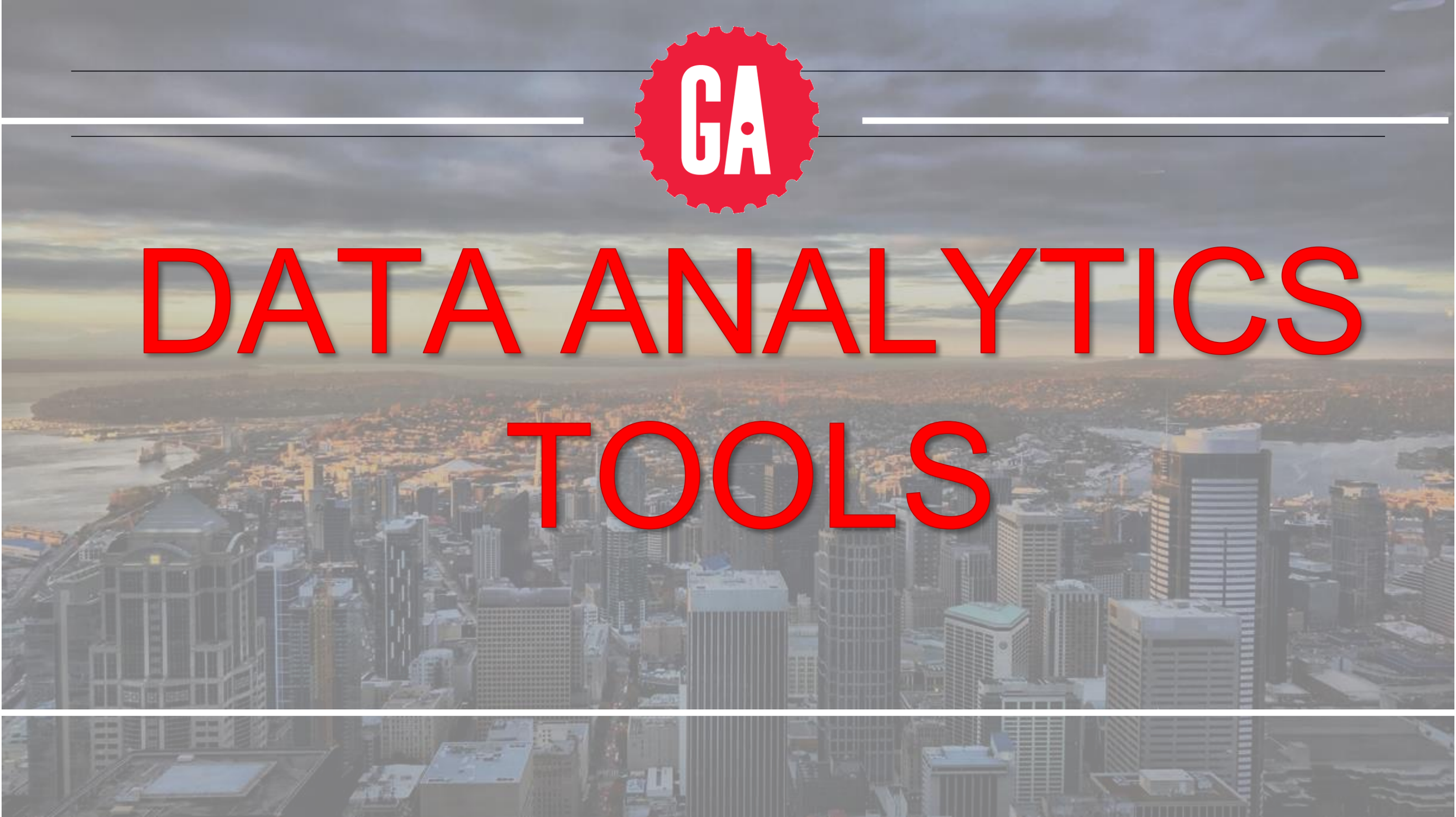
Data Analytics

Detailed examination of the elements or structure of something, typically as a basis for discussion or interpretation.

Data Analytics

Professional whose focus of analysis and problem solving relates to data, types of data and relationships among data elements within business or IT systems.

So what does all of this mean to you?


GA
DATA ANALYTICS
TOOLS

Data Tools

Excel – Preparing, relating, normalizing and adding value to data

PostGreSQL/PGAdmin – Obtaining, Filtering, Aggregating and adding value to data from a database

Tableau – Connecting to data, telling stories with data, Tableau workflow

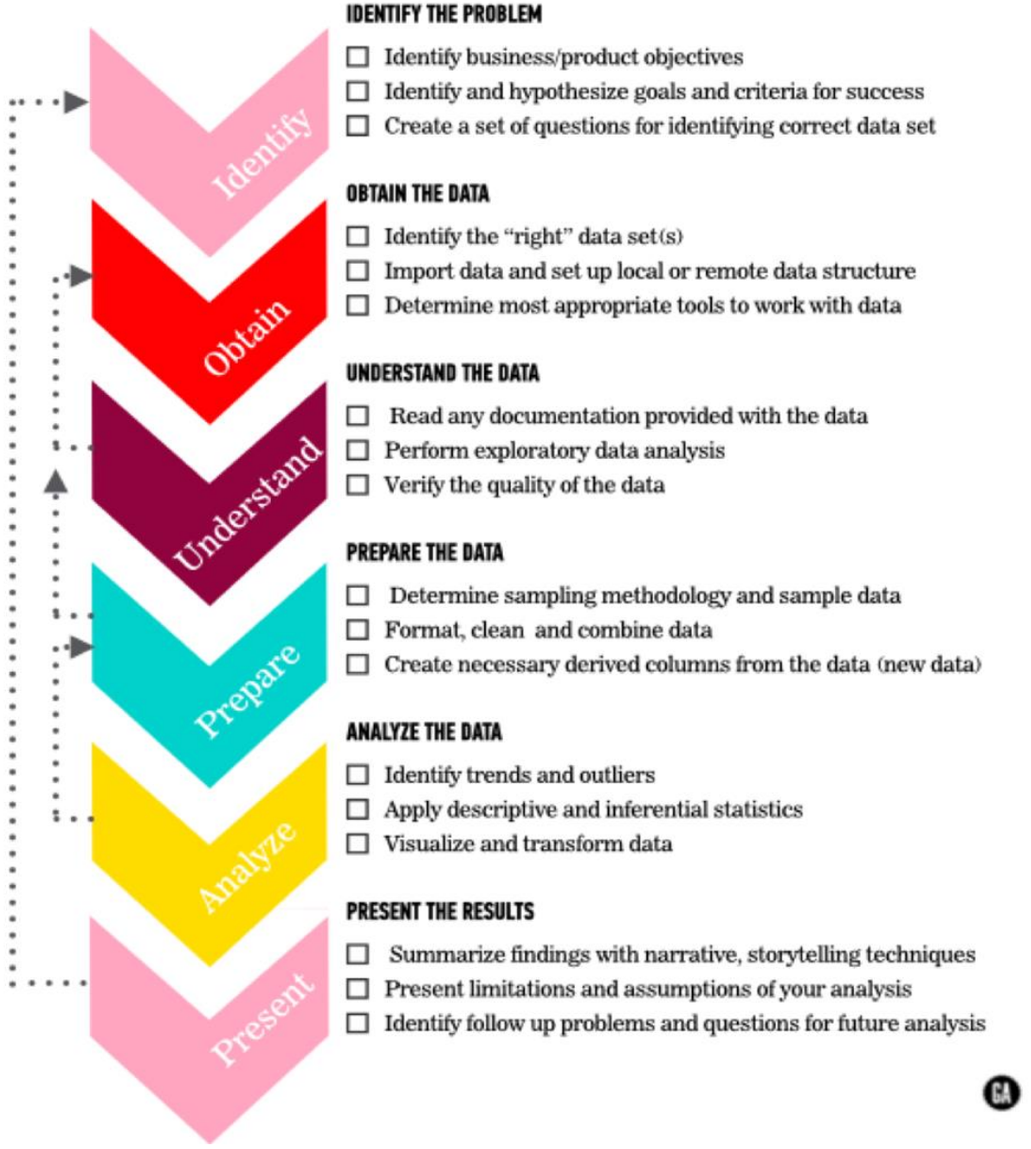
Data Tools

AirBnB

Mozilla Firefox

Capstone Project

Data Tools



IDENTIFY THE PROBLEM

- ☐ Identify business/product objectives
- ☐ Identify and hypothesize goals and criteria for success
- ☐ Create a set of questions for identifying correct data set

OBTAIN THE DATA

- ☐ Identify the "right" data set(s)
- ☐ Import data and set up local or remote data structure
- ☐ Determine most appropriate tools to work with data

UNDERSTAND THE DATA

- ☐ Read any documentation provided with the data
- ☐ Perform exploratory data analysis
- ☐ Verify the quality of the data

PREPARE THE DATA

- ☐ Determine sampling methodology and sample data
- ☐ Format, clean and combine data
- ☐ Create necessary derived columns from the data (new data)

ANALYZE THE DATA

- ☐ Identify trends and outliers
- ☐ Apply descriptive and inferential statistics
- ☐ Visualize and transform data

PRESENT THE RESULTS

- ☐ Summarize findings with narrative, storytelling techniques
- ☐ Present limitations and assumptions of your analysis
- ☐ Identify follow up problems and questions for future analysis

DATA ANALYTICS

Pre-work group Exercise/review

DEMO: BUSINESS SCENARIO

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

INDEPENDENT PRACTICE: CLEANING OUR DATASET

- 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.

INDEPENDENT PRACTICE: CLEANING OUR DATASET

- 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.

ACTIVITY: CLEANING OUR DATASET

EXERCISE

DIRECTIONS

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

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

TOOLS
P.E.M.D.A.S.
Part of whole
Adding decimals and finding difference from whole
Find replace (cleaning data)

DELIVERABLE

Complete BASE or STRETCH tab in 2014_acs_select_WA.xlsx

GUIDED PRACTICE: REVIEWING OUR SOLUTIONS

GUIDED PRACTICE: 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.

GUIDED PRACTICE: 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.

GUIDED PRACTICE: 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.

GUIDED PRACTICE: 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.

GUIDED PRACTICE: 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.

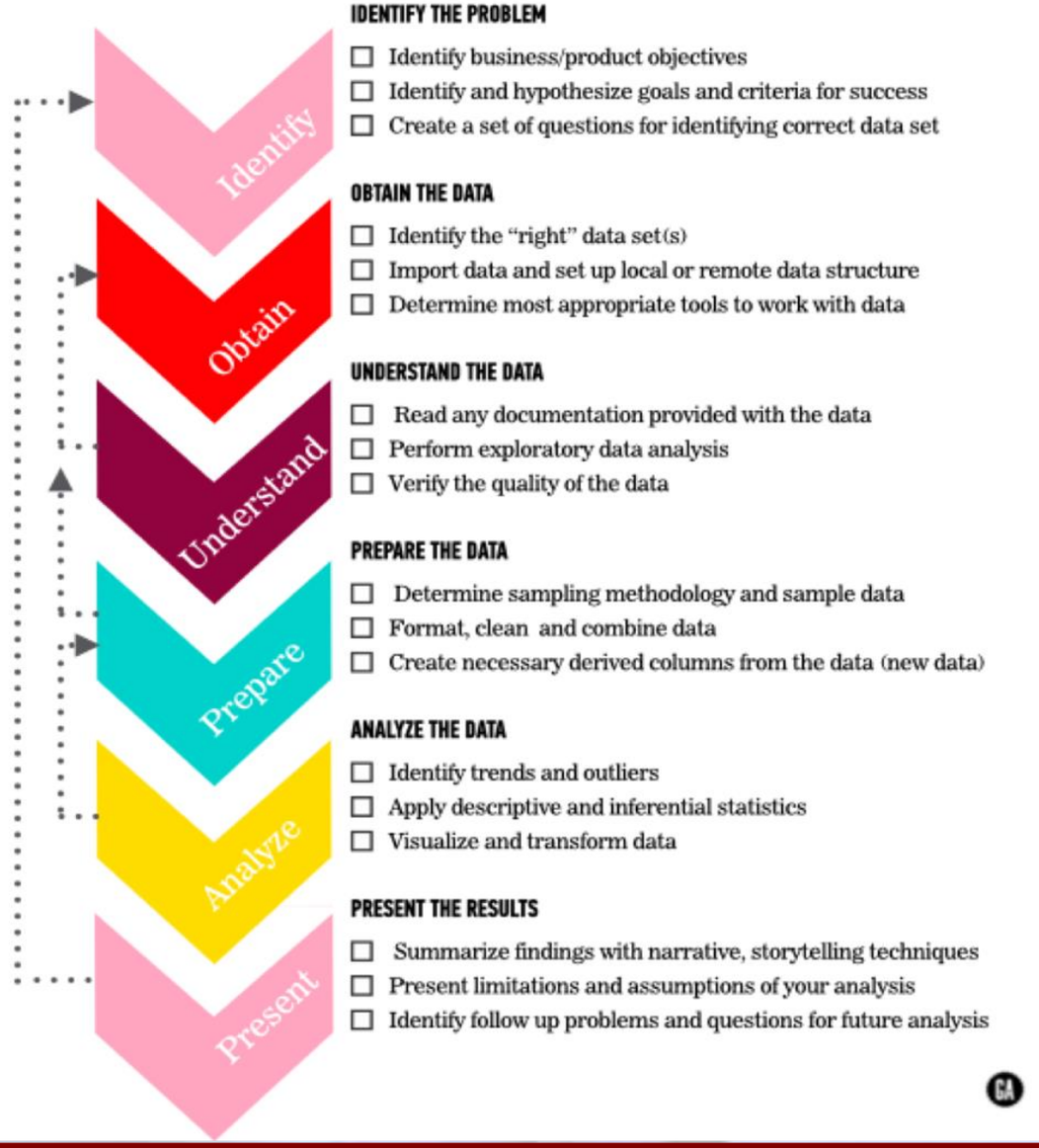
GUIDED PRACTICE: REVIEW SOLUTIONS

- 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.

4 essential tools

Workflow,6w's,Source, Summary

Data Tools



IDENTIFY THE PROBLEM

- ☐ Identify business/product objectives
- ☐ Identify and hypothesize goals and criteria for success
- ☐ Create a set of questions for identifying correct data set

OBTAIN THE DATA

- ☐ Identify the "right" data set(s)
- ☐ Import data and set up local or remote data structure
- ☐ Determine most appropriate tools to work with data

UNDERSTAND THE DATA

- ☐ Read any documentation provided with the data
- ☐ Perform exploratory data analysis
- ☐ Verify the quality of the data

PREPARE THE DATA

- ☐ Determine sampling methodology and sample data
- ☐ Format, clean and combine data
- ☐ Create necessary derived columns from the data (new data)

ANALYZE THE DATA

- ☐ Identify trends and outliers
- ☐ Apply descriptive and inferential statistics
- ☐ Visualize and transform data

PRESENT THE RESULTS

- ☐ Summarize findings with narrative, storytelling techniques
- ☐ Present limitations and assumptions of your analysis
- ☐ Identify follow up problems and questions for future analysis

Q & A

“Data! Data! Data!
I can’t make bricks without clay. “

-Sir Arthur Conan Doyle, Author

CONCLUSION

- Just cleaning data and creating percentages can add immense value to a dataset. Prepping data is a major part of data analytics.
- In the next lesson, we will learn about some very helpful Excel features: `LOOKUP` functions, `INDEX` and `MATCH`.

PROJECT

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…

PROJECT - Deliverable

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.

EXIT TICKET

CLASS : FUNDAMENTALS OF DATA AND EXCEL (pre-work review)

QUESTION: What is a Data Analyst?

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”