

Hello!



# Intro to Data Analytics



# Craig Fryar

Partner, CDO, Data Divers; COO, Fundify



- Global BI and Data Executive
- Data Analytics Instructor, Courseware Designer
- Enterprise, private and government, C-level execs
- Product Board Advisor for multiple startups
- Gamer, Advanced PADI Diver



<https://www.linkedin.com/in/craigfryar>



# Welcome to General Assembly!

We empower people to pursue work they **love** through education in business, coding, data, and design.





**“The world is one big data  
problem.”**

Andrew McAfee, MIT scientist

# What You'll Be Saying...



I can identify ways to strategically use data in my role.

I can use data to make informed recommendations.

I can follow a data workflow to go from identifying a hypothesis to visualizing insights.

I can build a predictive model.

# Where We're Headed

## What We'll Cover

- What's data anyway?
- What does a data analyst do?
- How do I use the Data Analysis Workflow to go from raw data to powerful insights?
- Where do I go from here to develop my skills?

# What to Keep in Mind

## So We're on the Same Page...

- Make sure you have a Google account!
- This is an intro class, so we're going to be working from a foundational level with more of a bird's-eye view of the field.
- There will be time at the end for Q&A.



Intro to Data Analytics

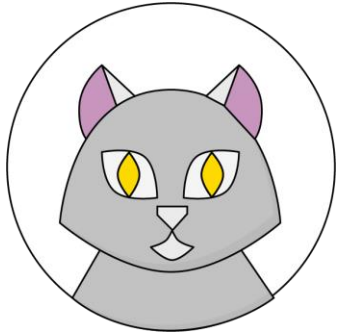
---

# What Is Data, Anyway?

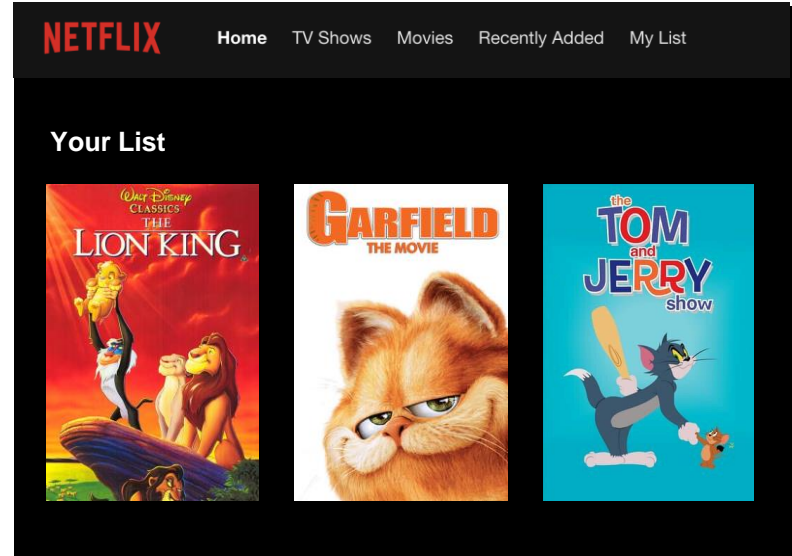


# What Exactly *Is* Data?

**Information** that exists in a variety of formats and sizes.



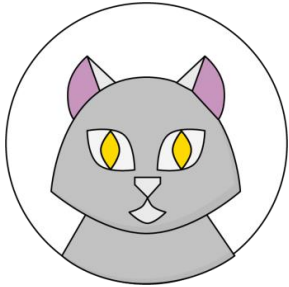
**Name:** Lala the Cat  
**Age:** 4  
**Parents:** Diana and Kevin  
**City:** Austin  
**Phone:** 416-555-MEOW  
**Instagram:** @lala\_lecat  
**Hobbies:** Nap, nip, and  
**Netflix**



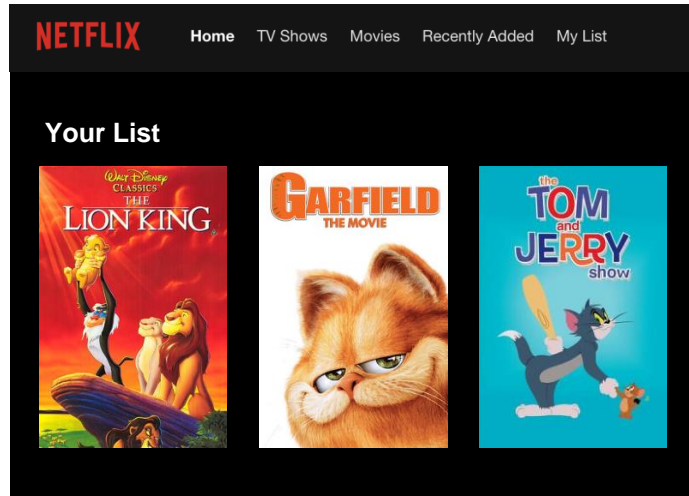


# What Is Data Analysis?

The process of **examining** data to draw **conclusions** about that information.



@lala\_lecat



← If you like this...

... you will also like this →



Intro to Data Analytics

---

# What Does a Data Analyst Do?



# What Do They Do?

**Data analysts** work to interpret data to bring **meaning** and **actionable insights**.



# Who Does It?

You don't need to be a data analyst to perform data analysis.

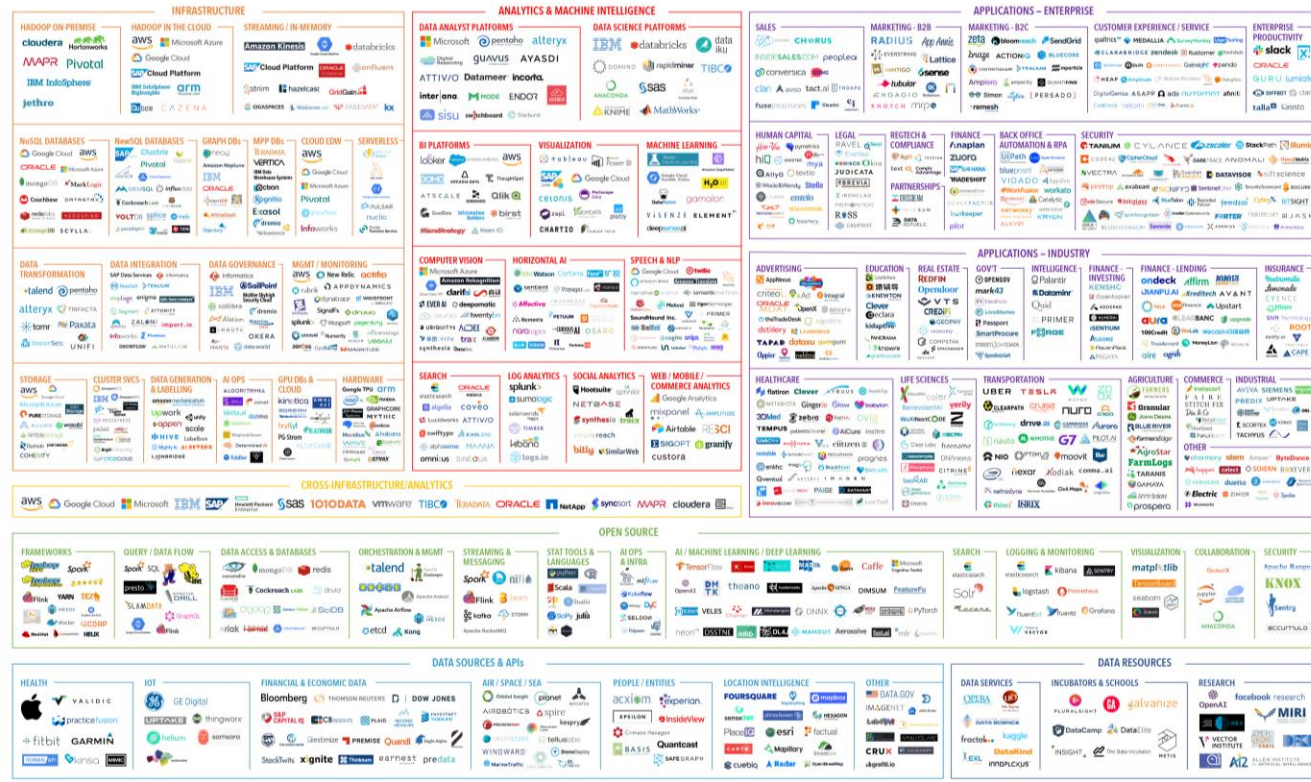


Do we have a product-market fit?

What type of hire will be most successful here?

Where should we be spending our marketing dollars?

# Data Analytics Landscape



# A Data Analyst's Toolkit

Excel



Basic data and analytics “scratchpads”

Sheets

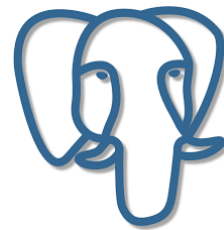


Tableau



Visual analysis,  
dashboard, and  
reporting platform

SQL



Data access and  
querying tool

# For Now: Google Sheets

You don't need to get fancy to gain powerful insights.

- Requires no setup (you only need a Google/Gmail account).
- Cross-platform.
- Free!



Intro to Data Analytics

---

# The Data Analysis Workflow





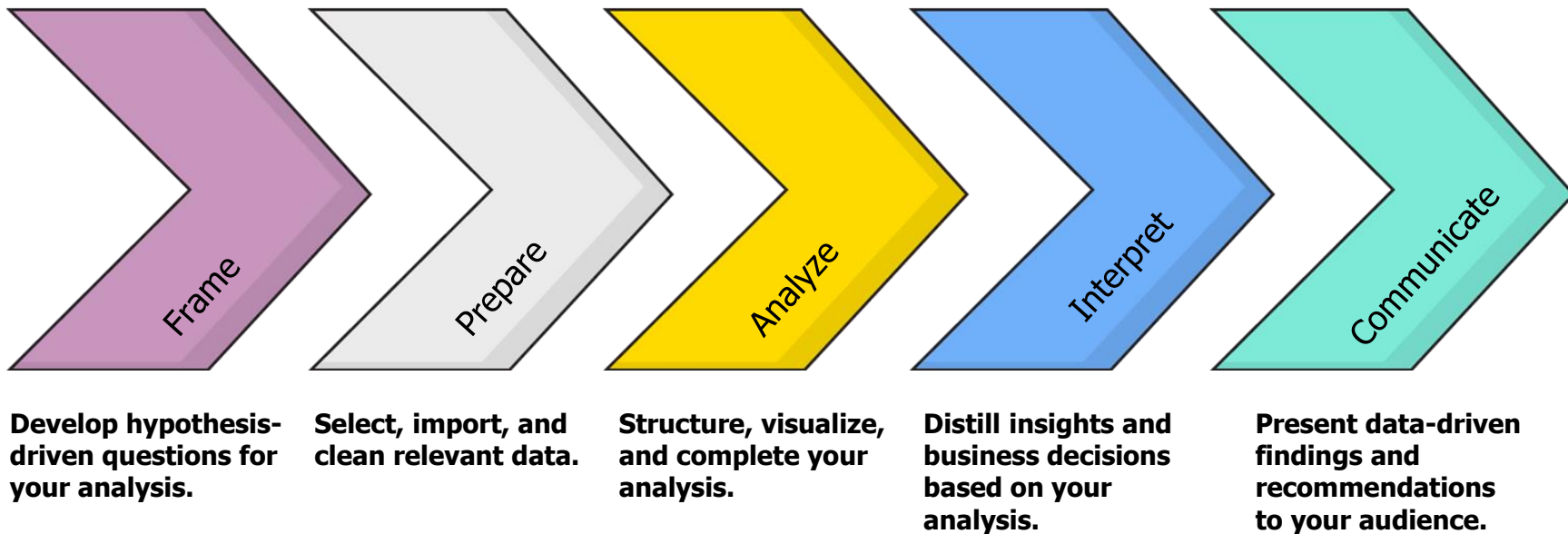
# Congrats – You Work at Netflix (*kinda*)!

## Your challenge:

Netflix wants to provide users with recommendations for what they should watch after finishing a show.



# GA's Data Analysis Workflow



Intro to Data Analytics

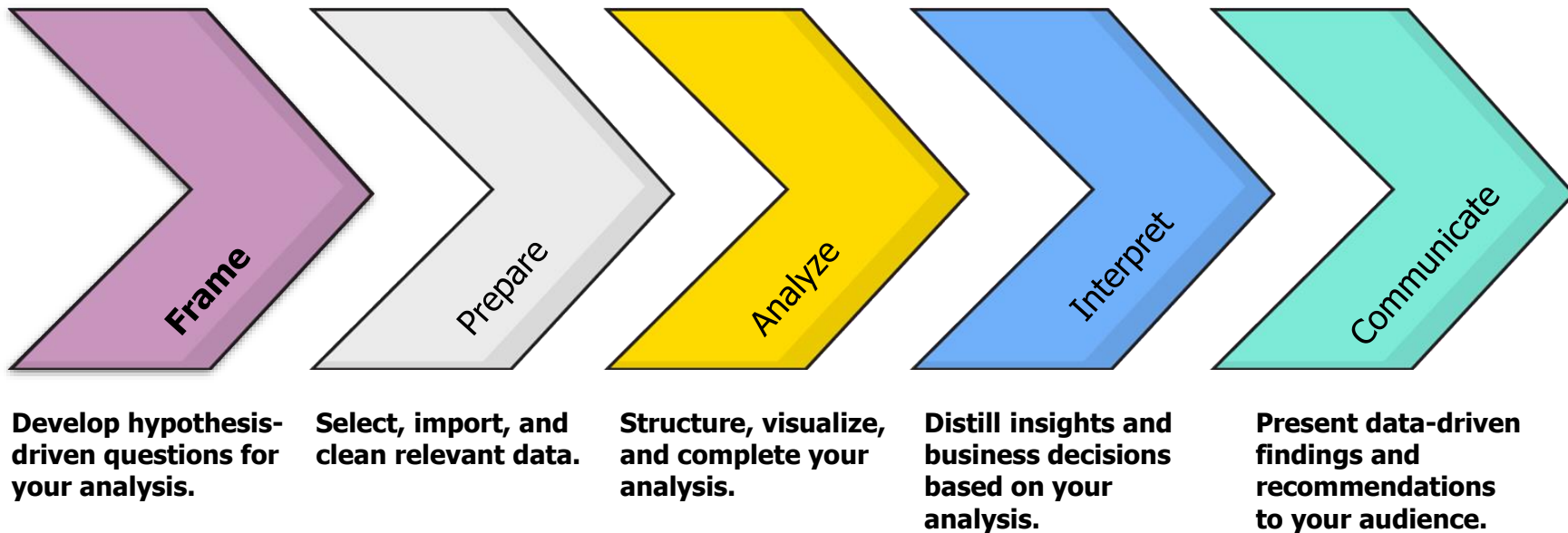
---

# Framing the Data



# GA's Data Analysis Workflow

You are here.






## Guided Walk-Through: Netflix Challenge



### **Your challenge:**

Netflix wants to provide users with recommendations for what they should watch after finishing a show.



How might we phrase our hypothesis-driven questions for analysis?

Intro to Data Analytics

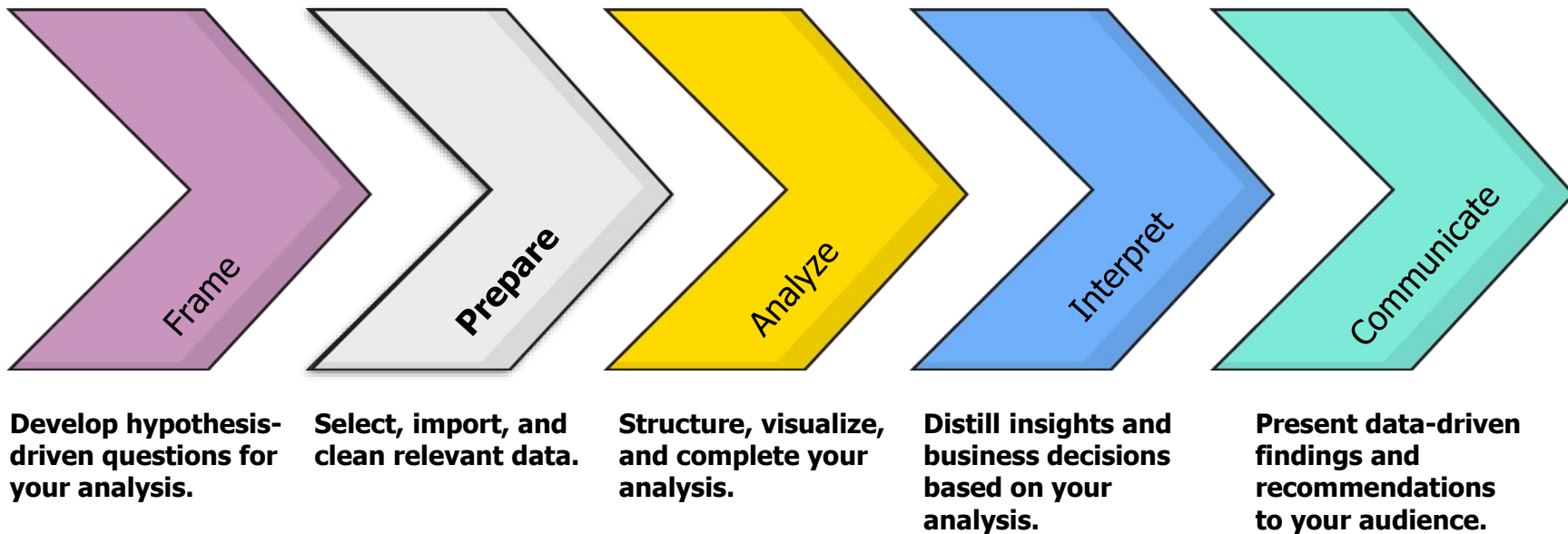
---

# Preparing the Data



# GA's Data Analysis Workflow

You are here.



# The Internet in Real-Time

How Quickly Data is Generated



Source: <https://visual.ly/community/infographic/how/internet-real-time>



# Obtaining Data

## Request Data



**Data  
Engineers**

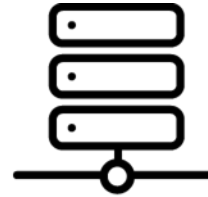


**IT  
Department**

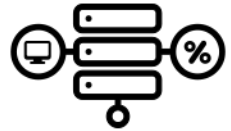
## Pull Data



**Reports**



**Databases**



**APIs**

# Obtaining Data

Request data



Data  
Engineers



IT  
Department

Pull data

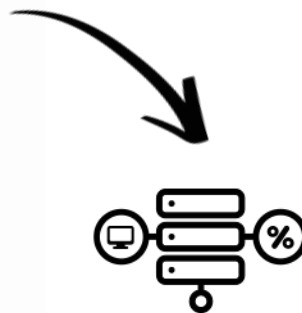


Reports



Databases

Quick  
definition:  
**API**



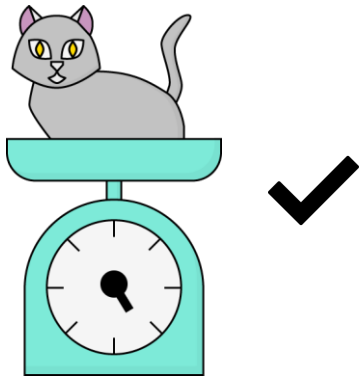
≠ IPAs



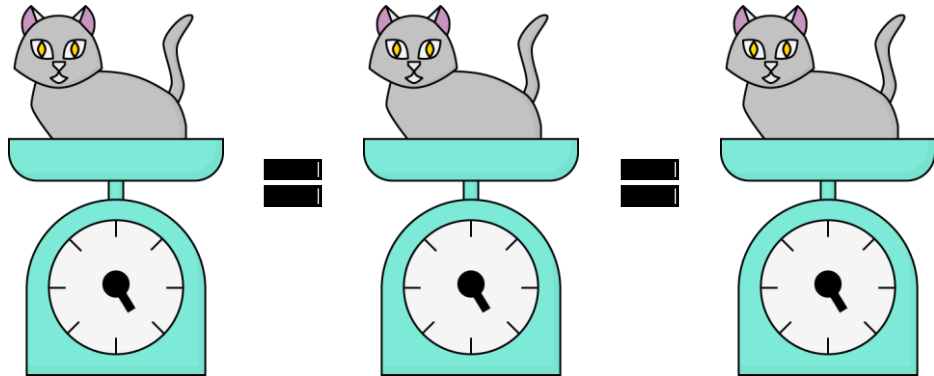
# Selecting Your Data

There are two main questions to keep in mind when you're deciding which data to use:

Is my data **valid**?

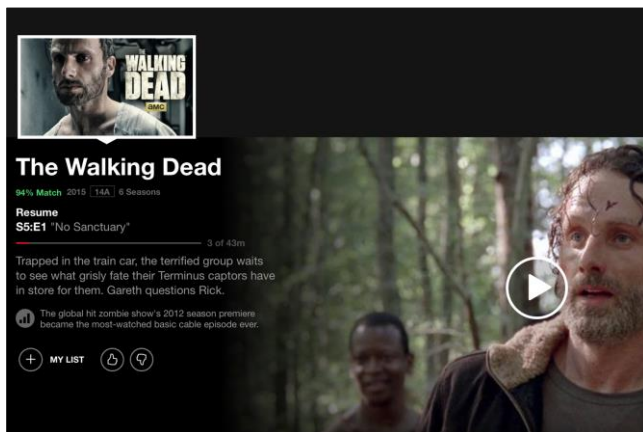


Is my data **reliable**?

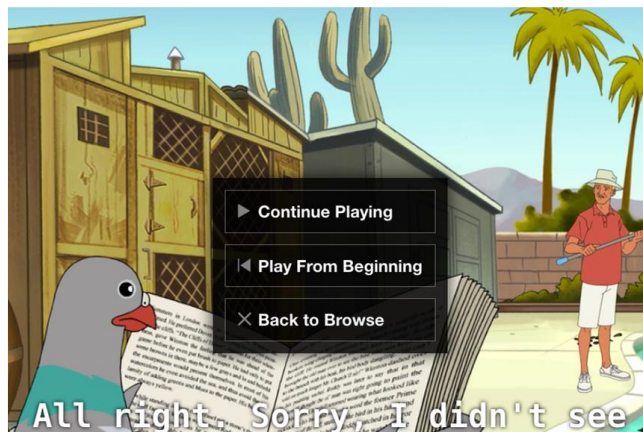




Someone at Netflix provides you with user viewing data. This data tracks what users are watching and for how long. Here's how Netflix gets that information:



Logging starts and stops



Still watching?



# Computers Out:

## Netflix Challenge: Prepare the Data



### Data Analytics Exercise: Netflix Raw Data

File Edit View Insert Format Data Tools Add-ons Help



100%



View only

	A	B	C	D	E	F
1	Subscriber ID	Show	Episodes Watched	Episode Length (m)		
2	SID_2000	Walking Dead	2.63	40		
3	SID_2000	Lost	0.60	45		
4	SID_2000	Weeds	0.27	30		
5	SID_2000	Mad Men	0.40	45		
6	SID_2000	Pony	0.07	30		
7	SID_2000	Lorax	0.03	30		
8	SID_1999	Walking Dead	3.20	40		
9	SID_1999	Lost	1.13	45		
10	SID_1999	Pony	0.17	30		
11	SID_1999	Weeds	0.13	30		
12	SID_1999	Lorax	0.10	30		
13	SID_1999	Mad Men	0.04	45		
14	SID_1998	Walking Dead	2.77	40		
15	SID_1998	Lost	0.82	45		
16	SID_1998	Weeds	0.37	30		
17	SID_1998	Mad Men	0.58	45		
18	SID_1998	Pony	0.10	30		
19	SID_1998	Lorax	0.03	30		
20	SID_1997	Walking Dead	2.68	40		

### 1) Open the data:

A-D - [ga.co/3b9JuWH](https://ga.co/3b9JuWH)

E-H - [ga.co/2SEvM84](https://ga.co/2SEvM84)

I-L - [ga.co/3dq7ZAK](https://ga.co/3dq7ZAK)

M-P - [ga.co/2W96ZuL](https://ga.co/2W96ZuL)

Q-T - [ga.co/2A5sXq4](https://ga.co/2A5sXq4)

U-Z - [ga.co/2U4vIVk](https://ga.co/2U4vIVk)

### 2) Make a Copy of the data set on your Drive.





# Computers Out: Netflix Challenge: Prepare the Data



## Step 1: Story of One Row:

describe the general nature of the data by examining one row (as if to someone unfamiliar with data analytics or the industry).

	A	B	C	D	E	F
1	Subscriber ID	Show	Episodes Watched	Episode Length (m)		
2	SID_2000	Walking Dead	2.63	40		
3	SID_2000	Lost	0.60	45		
4	SID_2000	Weeds	0.27	30		
5	SID_2000	Mad Men	0.40	45		
6	SID_2000	Pony	0.07	30		
7	SID_2000	Lorax	0.03	30		
8	SID_1999	Walking Dead	3.20	40		
9	SID_1999	Lost	1.13	45		
10	SID_1999	Pony	0.17	30		



## Step 2: Data Wrangling:

add new column to calculate "Total Minutes."

	A	B	C	D	E	F
1	Subscriber ID	Show	Episodes Watched	Episode Length (m)	Total Minutes Watched	
2	SID_2000	Walking Dead	2.63	40	105.2	
3	SID_2000	Lost	0.60	45	27.0	
4	SID_2000	Weeds	0.27	30	8.1	
5	SID_2000	Mad Men	0.40	45	18.0	
6	SID_2000	Pony	0.07	30	2.1	
7	SID_2000	Lorax	0.03	30	0.9	
8	SID_1999	Walking Dead	3.20	40	128.0	
9	SID_1999	Lost	1.13	45	50.9	
10	SID_1999	Pony	0.17	30	5.1	





# Computers Out: Netflix Challenge: Prepare the Data



**Step 3: Data Prep:** pivot the data by Subscriber ID to show Total Minutes Watched for each show by Subscriber.

Subscriber ID	Lorax	Lost	Mad Men	Pony	Walking Dead	Weeds
SID_2000	0.9	27.0	18.0	2.1	105.2	8.1
SID_1999	3.0	50.9	1.8	5.1	128.0	3.9
SID_1998	0.9	36.9	26.1	3.0	110.8	11.1
SID_1997	6.9	34.2	34.2	5.1	107.2	11.1
SID_1996	5.1	57.2	16.2	3.0	128.8	9.0
SID_1995	0.9	64.8	16.2	0.9	132.0	8.1
SID_1994	3.9	54.0	5.9	3.0	126.8	6.0
SID_1993	6.0	71.1	1.8	3.9	142.8	5.1
SID_1992	3.9	58.1	14.0	3.0	127.2	6.9
SID_1991	0.9	32.9	22.1	2.1	105.2	6.9
SID_1990	0.9	50.0	1.8	2.1	120.8	3.0
SID_1989	0.9	61.2	12.2	2.1	130.0	5.1
SID_1988	6.9	45.0	4.1	6.0	116.0	0.9
SID_1987	3.0	72.0	0.0	2.1	136.8	3.0
SID_1986	8.1	45.0	9.9	6.0	118.0	3.9
SID_1985	3.9	45.9	0.0	6.0	117.2	3.0
SID_1984	5.1	54.0	19.8	6.9	121.2	9.0
SID_1983	3.9	54.9	0.0	2.1	119.2	0.0
SID_1982	2.1	45.0	18.0	5.1	108.8	6.0

Intro to Data Analytics

---

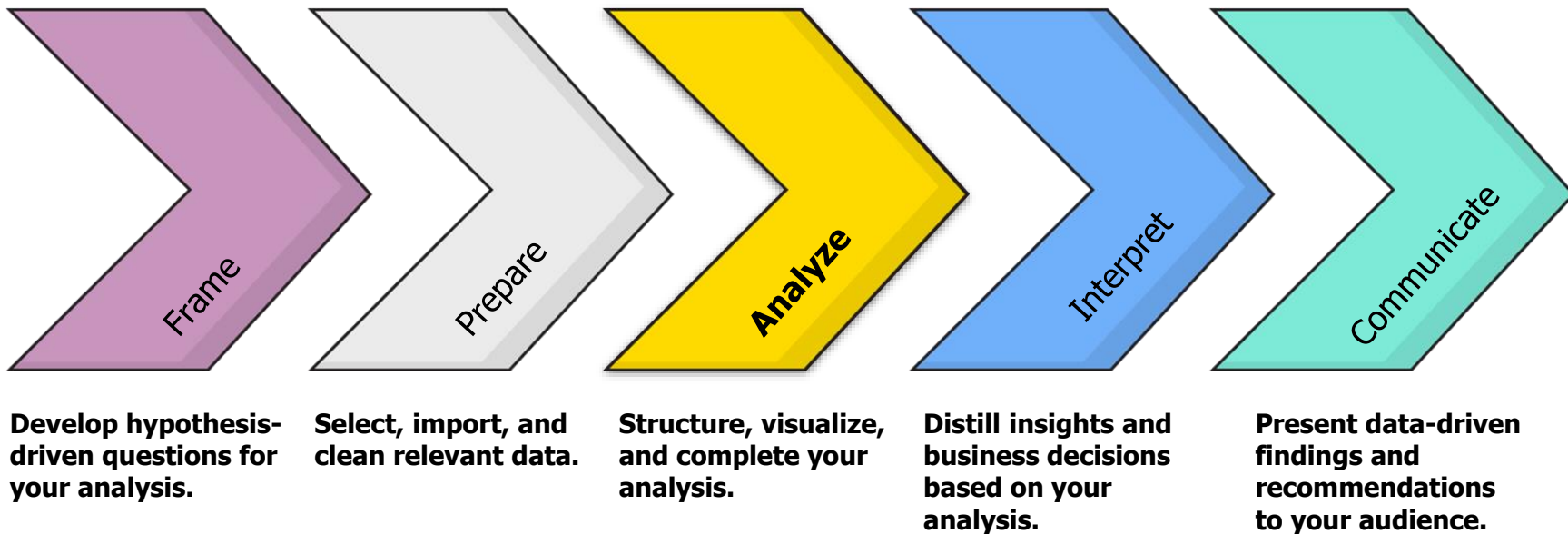
# Analyzing the Data





# GA's Data Analysis Workflow

You are here.



# Let's Talk Math: What Are Statistics?

## Descriptive Statistics

Summarized information about a collection of data, also called a data set.

U.S. population: 323 million  
U.S. median age: 37.6

## Inferential Statistics

Models that let us draw conclusions about a population using sample data.

Election polls

## Predictive Statistics

Models that let us anticipate how members of a population are likely to behave.

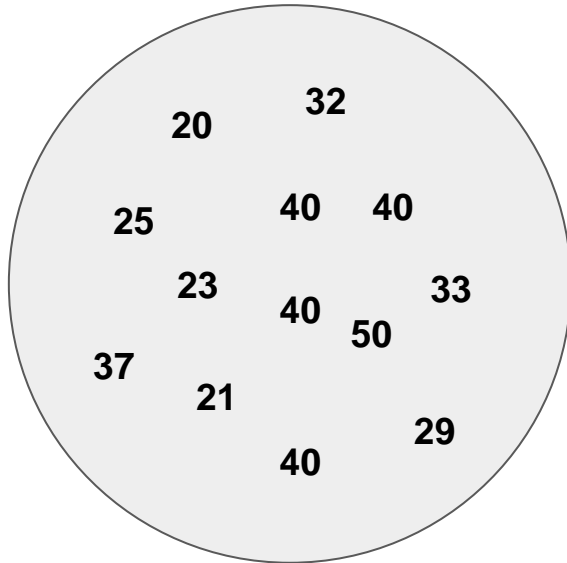
Weather forecasts  
Recommendation engines

**This is what  
we're doing!**



# Let's Talk Math: Descriptive Statistics

Information that is **most** indicative of a large set of data → average



**Mean = 33.08**

**Sum:**

$20 + 25 + 32 + 40 + 40 + 23$   
 $+ 33 + 50 + 21 + 40 + 29$   
 $+ 37 + 40 = 430$

**Total observation: 13**

**Mean =  $430 / 13$**

**Mode = 40**

**Most repeated  
observation: 40**

**Median**

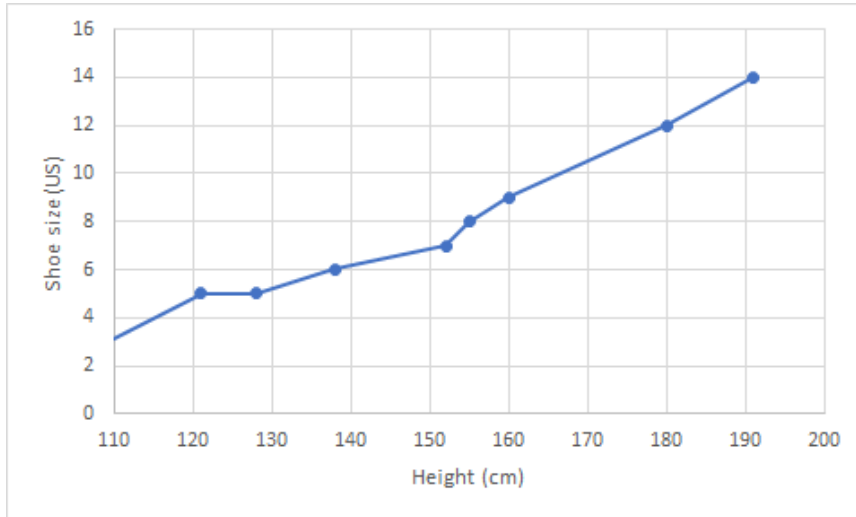
20 21 23 25 29 32 **33** 37 40 40 40 40 50

# Let's Talk Math: Correlations

“Co” (together) + “relation” = The relationship between two variables.

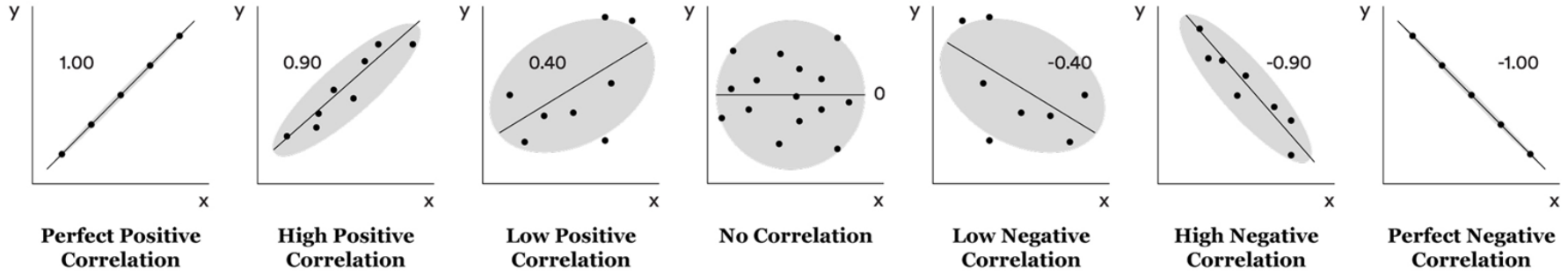
**Positive:** When values are increasing together.

**Negative:** When one value is increasing and the other is decreasing.



Height	Shoe Size
109	3
121	5
138	6
155	8
160	9
180	12
191	14

# Let's Talk Math: Correlations (Cont.)



# Descriptive Statistics

## Definition

Summarized information about a collection of data, also called a data set.

## Answer These Questions

- What are the most watched shows?
- What shows have the highest average viewing time?





# Computers Out: Netflix Challenge: Analyze the Data



Let's examine our **descriptive statistics**.

Open up the Google Sheet containing our Netflix data and follow along.

Here's the  Google Sheet links:

A-D - [ga.co/3b9JuWH](https://ga.co/3b9JuWH)

E-H - [ga.co/2SEvM84](https://ga.co/2SEvM84)

I-L - [ga.co/3dq7ZAK](https://ga.co/3dq7ZAK)

M-P - [ga.co/2W96ZuL](https://ga.co/2W96ZuL)

Q-T - [ga.co/2A5sXq4](https://ga.co/2A5sXq4)

U-Z - [ga.co/2U4vIVk](https://ga.co/2U4vIVk)

Copy of Exercise: Netflix Raw Data

File Edit View Insert Format Data Tools Add-ons Help

fx

	A	B	C	D	E	F	G
1	Subscriber ID	Show	Episodes Watched	Episode Length (m)			
2	2000	Walking Dead	2.63	40			
3	2000	Lost	0.60	45			
4	2000	Weeds	0.27	30			
5	2000	Mad Men	0.40	45			
6	2000	Game of Thrones	0.27	30			



# Computers Out: Netflix Challenge: Analyze the Data



Here's what we see.

What is this data showing us?

Total Average Minutes Watched	
Lorax	3.72
Lost	52.11
Mad Men	13.15
Pony	4.01
Walking Dead	120.23
Weeds	5.94
Grand Total	33.19

	MIN of Total Minutes	MAX of Total Minutes
Lorax	0	9.9
Lost	0.9	94.05
Mad Men	0	52.2
Pony	0	14.1
Walking Dead	78	153.2
Weeds	0	17.1
Grand Total	0	153.2



# Correlation

## Definition

The relationship between two variables.

## Answer the Question

How common it is that people view both Show A *and* Show B?



# Welcome to the Matrix

A **correlation matrix** allows us to understand the relationships between multiple variables; in this case, between time spent watching shows.

	Lorax	Lost	Mad Men	Pony	Walking Dead	Weeds
Lorax						
Lost						
Mad Men						
Pony						
Walking Dead						
Weeds						



Let's examine **correlations** in our data set.

Open up the Google Sheet containing our Netflix data and follow along.



## Computers Out: Netflix Challenge: Analyze the Data (Cont.)



Here's what we see.

What are we observing in this data?

	Lorax	Lost	Mad Men	Pony	Walking Dead	Weeds
Lorax	1					
Lost	-0.00151	1				
Mad Men	-0.02271	-0.12902	1			
Pony	0.57499	-0.01994	-0.02349	1		
Walking Dead	-0.01166	0.95334	-0.23833	-0.01760	1	
Weeds	-0.01936	-0.12606	0.90489	-0.02798	-0.21414	1

Intro to Data Analytics

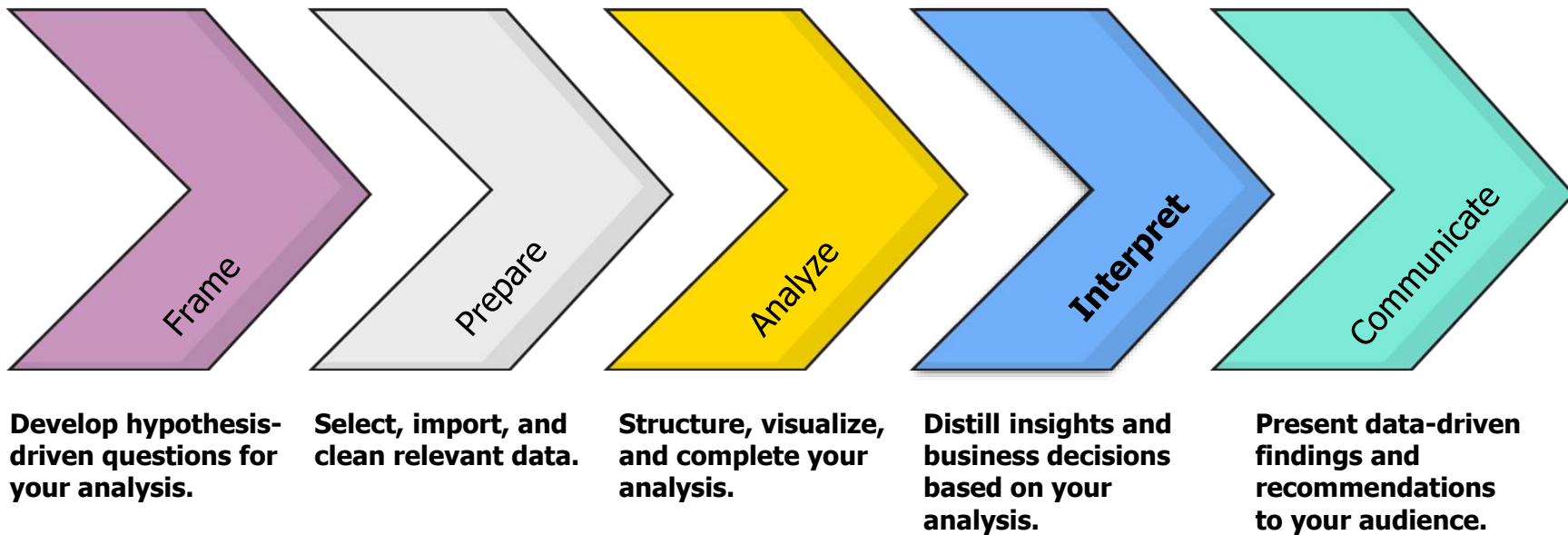


# Interpreting the Data



# GA's Data Analysis Workflow

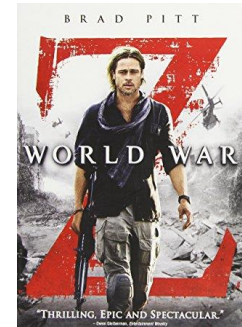
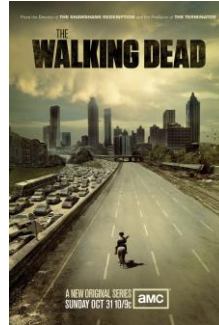
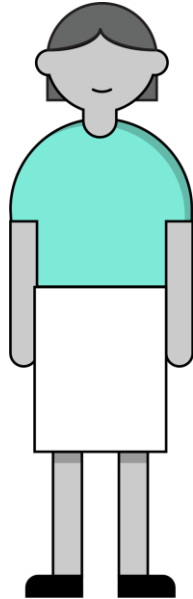
You are here.



# Interpreting the Data

Intuitively

...





## Discussion:

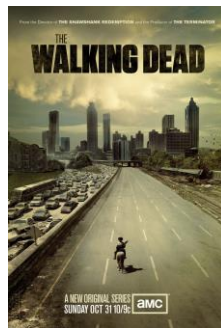
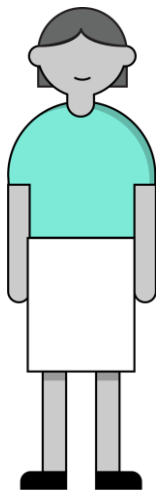
# Netflix Challenge: Interpret the Data



Based on what we know about users who watch The Walking Dead...

- What would you recommend next?
- What would you not recommend?

**Based on  
our data...**







## Discussion:

# Netflix Challenge: Interpret the Data (Cont.)



Based on this model, how should Netflix decide if they should spend \$10,000 on the distribution rights to bring Weeds to Canada?



	Lorax	Lost	Mad Men	Pony	Walking Dead	Weeds
Lorax	1					
Lost	-0.00151	1				
Mad Men	-0.02271	-0.12902	1			
Pony	0.57499	-0.01994	-0.02349	1		
Walking Dead	-0.01166	0.95334	-0.23833	-0.01760	1	
Weeds	-0.01936	-0.12606	0.90489	-0.02798	-0.21414	1



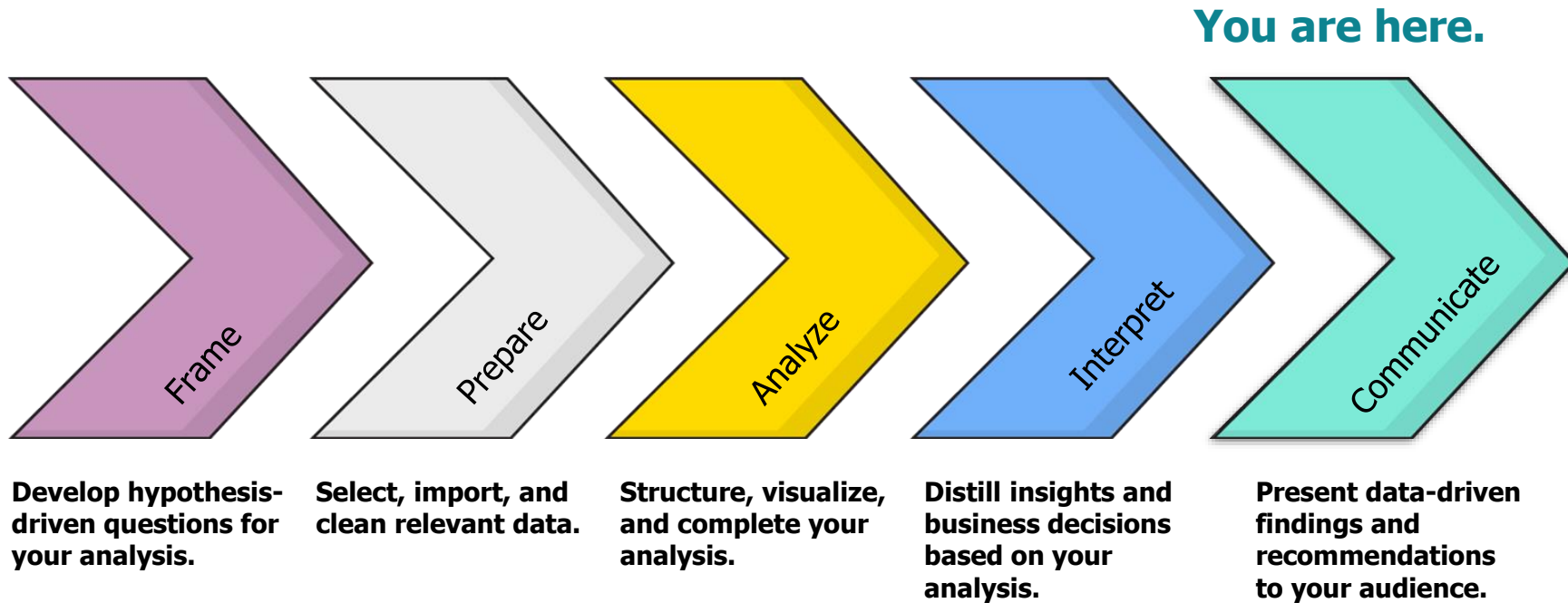
Intro to Data Analytics

---

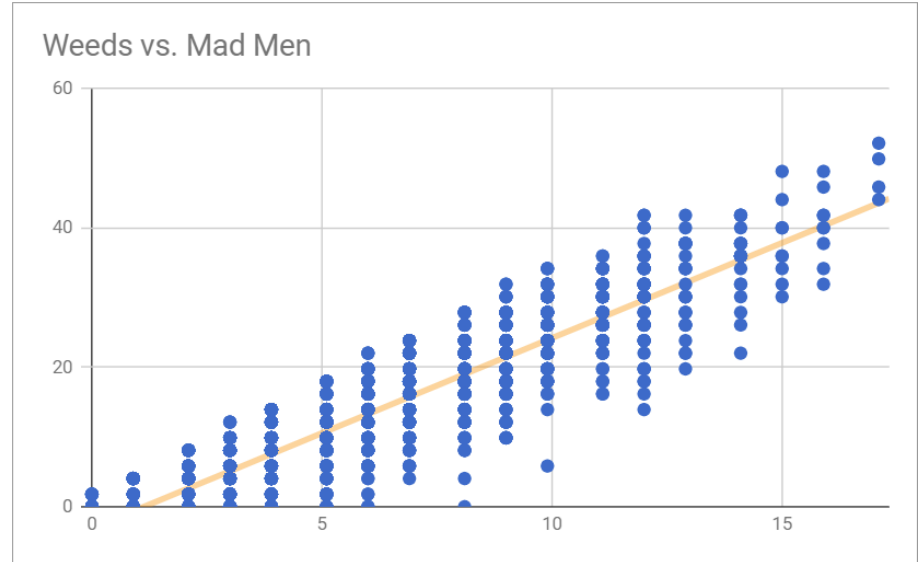
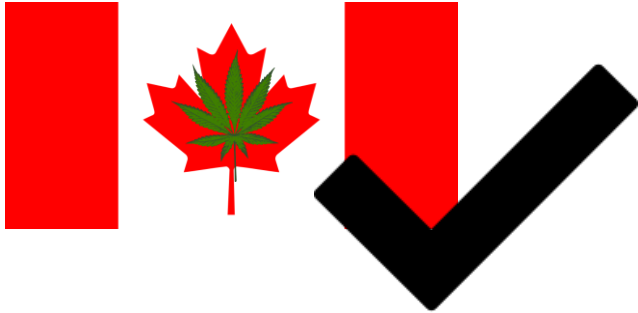
# Communicating the Data



# GA: Data Workflow



# Data Visualization



Intro to Data Analytics

---

# What's Next?



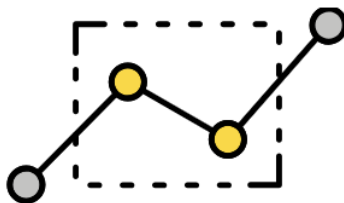
# Data at GA

## Data Analytics



Part time  
On campus  
Online

## Data Science



Full time  
Part time  
On campus  
Online

# Want to Learn More?

**GENERAL ASSEMBLY**

On Campus \* Online \* Topics \* Locations \* For Companies \* About \* **Find Your Course** Sign In

VIEW ALL DATA COURSES

# DATA ANALYTICS

10-WEEK PART-TIME OR 1-WEEK ACCELERATED COURSE

Request Syllabus **Apply Now**

**Overview**

Learning Support

Curriculum

Instructors

## The Big Picture

Confidently make — and defend — critical decisions using the results of your data analysis.

## Skills, Tools, & Strategies

16

# Upcoming Data Analytics Courses:

Jul 27 – Aug 1

Mon - Sat: 9:00am - 5:00pm CDT

Online

[Apply](#)

Aug 4 – Oct 8

Tue & Thu: 6:00pm - 8:00pm CDT

Online

[Apply](#)

Aug 10 – Oct 19

Except: Aug 31

Mon & Wed: 12:00pm - 3:00pm CDT

Online

[Apply](#)

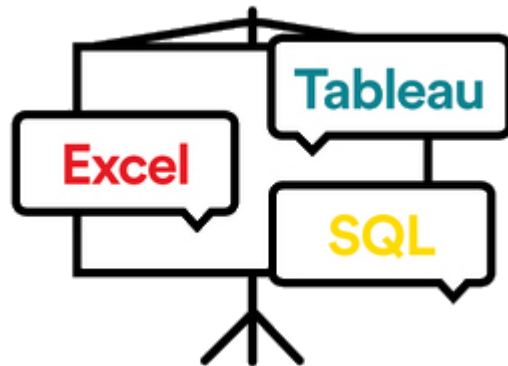
Aug 10 – Oct 19

Except: Sep 7

Mon & Wed: 5:00pm - 7:00pm CDT

Online

[Apply](#)



Apply Now

Request Syllabus



# EXCEL AT EXCEL: GETTING DOWN THE BASICS BOOTCAMP

## About this workshop

We will be going over how to:

- Terminology and Navigation
- Formatting
- Data Organization
- Conditional Formatting
- Hiding and Grouping
- Formulas
- Functions
- Basic Math Functions
- Reference Functions
- Text Functions
- Filling
- Defined Name
- PivotTables
- Chart
- Trendline
- Secondary Axis



10

Monday, 10 August

📍 Online



3

Thursday, 3 September

📍 Online



24

Thursday, 24 September

📍 Online



# SQL BOOTCAMP REMOTE (ONLINE)

## About this workshop

Imagine that you have a spreadsheet containing every data point about your customers — their preferences, their web browsing behavior, the products they've bought. What would you do with all that information? When you know SQL, you can slice and dice that data any way you want, providing critical insights that drive business strategy.

SQL provides powerful but reasonably simple tools for data analysis and handling. This bootcamp will take absolute beginners through the basics of SQL to an ability to write queries with confidence — and start using the tool immediately.

**Due to pre-work requirements, students must sign up at least 24 hours before the course start time.**

## Takeaways

- Understand SQL vocabulary such as tables, schema, functions and JOINS.
- Understand the differences between different data management systems for SQL databases.
- Create basic SQL queries by applying Boolean logic, sorting functions, and commenting.
- Apply SQL aggregate functions [MIN, MAX, SUM, AVG, COUNT].
- Use advanced SQL commands [OR, GROUP BY, HAVING] and conditional operators [=, !=, >, <, IN, NOT IN, and BETWEEN] to filter data.



Tuesday, 4 August

📍 Online



Monday, 17 August

📍 Online



Thursday, 27 August

📍 Online



# SQL BOOTCAMPS: INTRO & INTERMEDIATE



pgAdmin 4

127.0.0.1:55465/browser/

pgAdmin 4 File Object Tools Help

Browser

- FTS Templates
- Foreign Tables
- Functions
- Materialized Views
- Sequences
- Tables (4)
  - counties
  - products
  - Columns (16)
  - Constraints
  - Indexes
  - Rules
  - Triggers
  - sales
    - Columns (17)
      - date
      - convenience\_store
      - store
      - county\_number
      - county
      - category
      - category\_name
      - vendor\_no
      - vendor
      - item
      - description
      - pack
      - liter\_size
      - state\_btl\_cost

Query Editor

```
1 SELECT store,  
2     CAST(MIN(total) AS MONEY) AS MIN_TOTAL,  
3     CAST(MAX(total) AS MONEY) AS MAX_TOTAL,  
4     CAST(SUM(total) AS MONEY) AS SUM_TOTAL,  
5     ROUND(AVG(total),2) AS AVG_TOTAL,  
6     CAST(AVG(total) AS MONEY) AS AVG_TOTAL_CASTED  
7 FROM sales  
8 GROUP BY store  
9 ORDER BY SUM_TOTAL DESC;  
10
```

Data Output Explain Messages Notifications Query History

	store integer	min_total money	max_total money	sum_total money	avg_total numeric	avg_total_casted money
1	2633	\$1.79	\$94,590.72	\$13,920,087.22	472.43	\$472.43
2	4829	\$6.00	\$16,422.00	\$11,942,399.97	462.40	\$462.40
3	3420	\$39.72	\$41,055.00	\$6,159,480.06	1106.23	\$1,106.23
4	3385	\$35.40	\$41,055.00	\$5,734,721.57	803.75	\$803.75
5	2512	\$1.58	\$45,343.20	\$5,665,143.70	235.67	\$235.67
6	3814	\$67.43	\$86,382.00	\$4,907,465.88	1825.69	\$1,825.69
7	3952	\$7.50	\$6,522.00	\$4,289,169.59	319.73	\$319.73
8	3354	\$39.72	\$21,144.00	\$3,308,625.56	623.56	\$623.56
9	2625	\$2.70	\$13,477.92	\$3,169,984.14	229.00	\$229.00

# INTRO TO TABLEAU: DATA DRIVEN INSIGHTS BOOTCAMP REMOTE (ONLINE)



## About this workshop

We will be going over how to:

- Connect to and bring real-world data into Tableau;
- Easily clean and prepare data for exploration and interactive visual analysis;
- Explore different data types and methods to drill up and down into data detail;
- Build a variety of analysis chart types, including maps, time-series, and forecasts;
- Create calculated fields from provided data extract;
- Work with dates, including discrete and continuous time aggregations;
- Apply one-click advanced analytics to extend the power of data visualizations;
- Create parameters to enable end-user what-if analysis scenarios;
- Combine visualizations into dashboards, along with data filters and interactivity;
- Explore visual finishing and formatting techniques;
- Present insights from vizzes and dashboard using Tableau StoryPoints;
- See how to create free portfolios of dashboards and StoryPoints with Tableau Public;
- Present opportunities for intermediate Tableau training and industry certification.

29

Wednesday, 29 July

📍 Online



24

Monday, 24 August

📍 Online



21

Monday, 21 September

📍 Online





**We value your feedback!**

*Please take 60 seconds to complete our survey.*

[ga.co/introto](https://ga.co/introto)



# Explore a New Field With On Demand

**Premium content.** Build in-demand skills with bite-sized lessons in digital marketing, UX design, data analysis, and more.

**12-month access.** Learn on your schedule 24/7 on our dynamic, mobile-friendly platform.

**Dedicated expert support.** Get helpful project guidance and feedback during 1:1 sessions with mentors.

**Industry-backed assessments.** Pinpoint strengths and highlight opportunities for improvement.

Learn more at  
[ga.co/OnDemand](https://ga.co/OnDemand)

# Questions?



