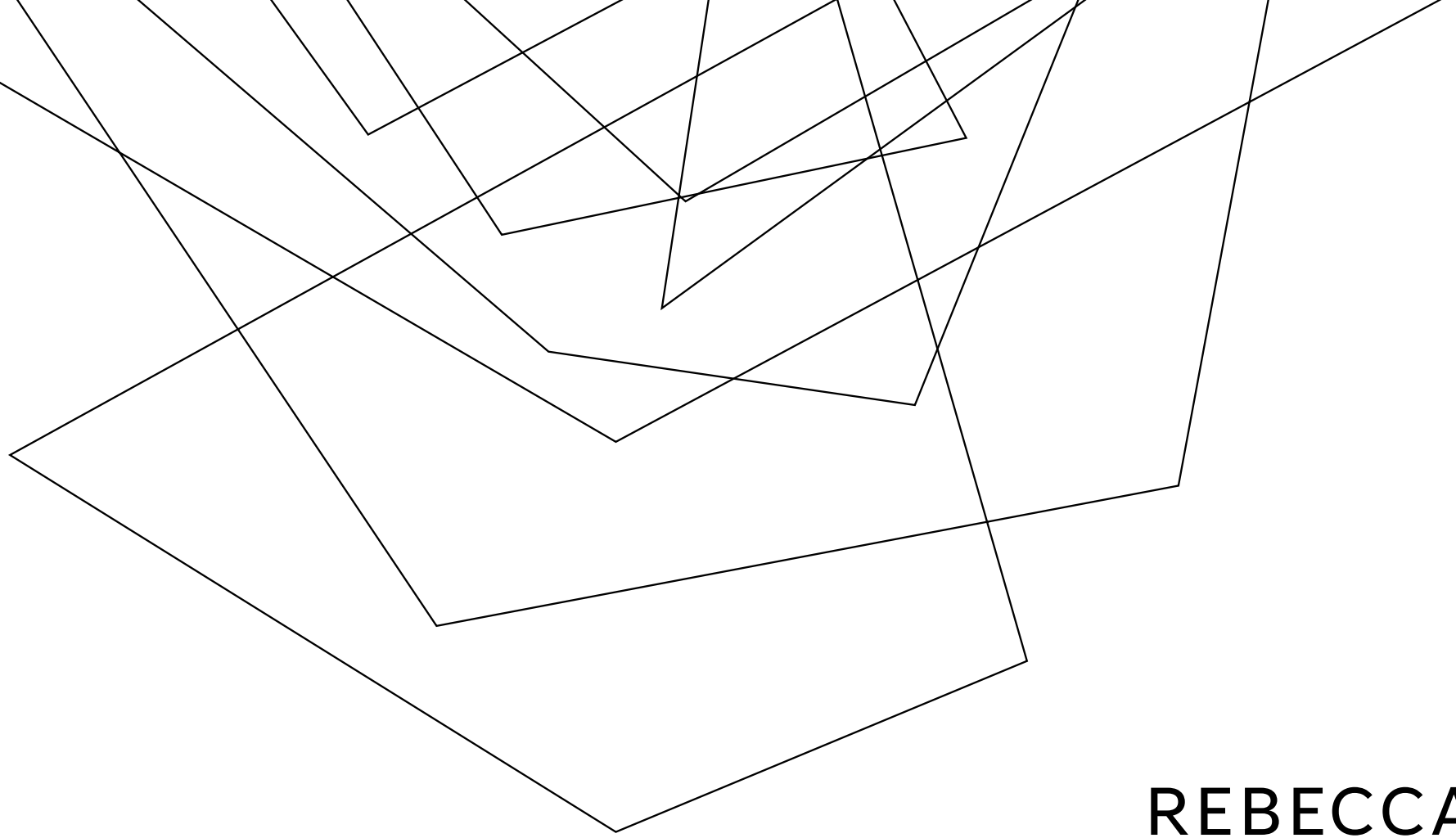


REBECCA MEYER

DATA ANALYST PORTFOLIO

RLWINTERISU@GMAIL.COM

SQL | PYTHON | TABLEAU

ABOUT ME

EDUCATION

- Bachelor’s Degree in Journalism and Mass Communication
- Certification in Data Analytics (in progress)

TOOLS

- Excel, Tableau, SQL, Python

BACKGROUND

With over 10 years of experience in email marketing, I've worked extensively with campaign data, customer behavior, A/B testing, and performance metrics. This experience inspired me to pivot and to pursue a career in analytics. I'm an aspiring data analyst with hands-on experience using tools like Excel, Tableau, SQL, and Python, and I’m actively building my portfolio through projects that reflect my curiosity and problem-solving mindset.

PROJECT OVERVIEW

	PROJECT	TOOLS
	VIDEO GAME POPULARITY DATA PROJECT	EXCEL, POWERPOINT
	PREPARING FOR INFLUENZA SEASON	EXCEL, TABLEAU
	ROCKBUSTER STEALTH DATA ANALYSIS PROJECT	SQL, POSTGRESQL, TABLEAU
	INSTACART GROCERY BASKET ANALYSIS	PYTHON, EXCEL
	LIBRARY FUNDING AND LITERACY RATES	PYTHON, EXCEL, TABLEAU

VIDEO GAME POPULARITY DATA PROJECT

Background:

GameCo's current understanding around sales for the various geographic regions assumes that sales stay the same over time.

Whether this is true or not, the marketing budget must be looked at to determine if the budget needs to be redistributed among the regions to maximize return on investment.

Objective:

- Analyze GameCo's sales from 1980 - 2016
- Determine if budget needs to be reallocated among regions

Tools:

- Excel
- PowerPoint

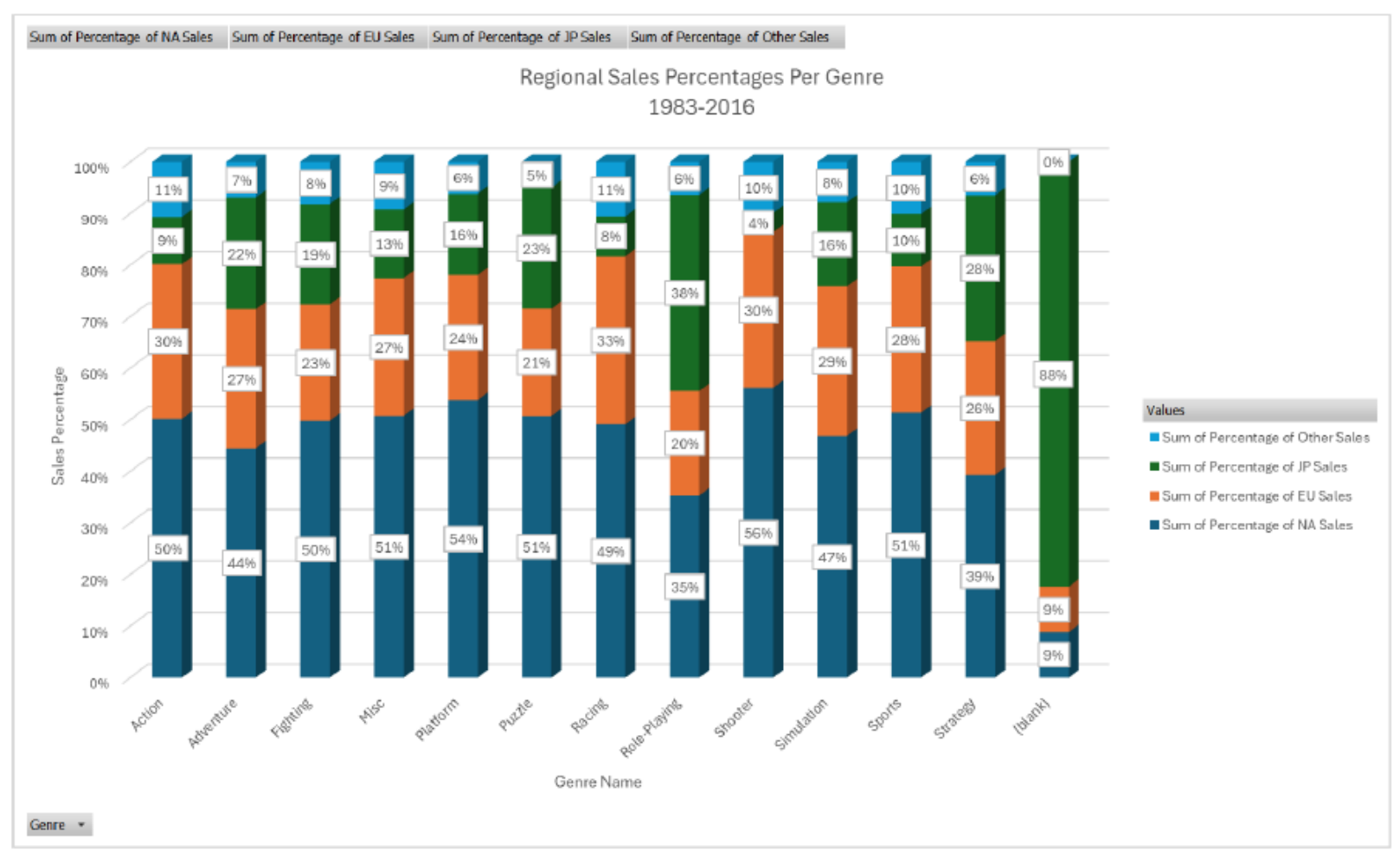
Key Questions:

1. Are certain types of games more popular than others?
2. Have any games decreased or increased in popularity over time?
3. How have sales figures varied between geographic regions over time?

Data Set:

Data pulled from the website VGChartz

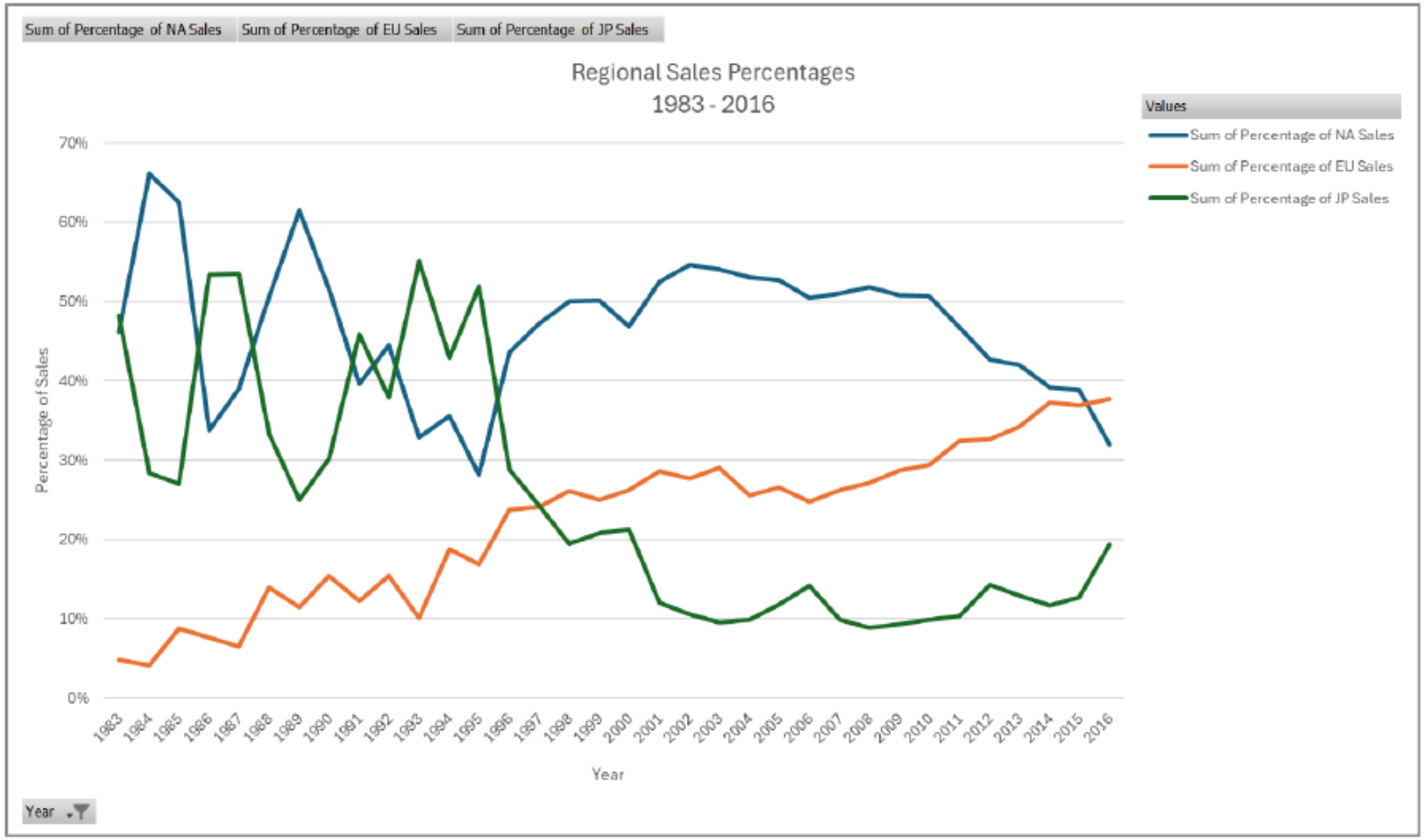
VIDEO GAME POPULARITY DATA PROJECT



100% stacked bar chart showing performance of regional sales in each genre.

- This chart breaks out how GameCo's regions performed for each genre.
- **NA dominates the market**, followed by EU.
- **JP is a key market for Role-Playing**, but shows little interest in Shooter games.

VIDEO GAME POPULARITY DATA PROJECT



Line chart showing regional sales percentages through years 1983-2016

- This chart shows how sales have **not remained the same throughout the years** for any region. There have been sharp increases and decreases in NA and JP sales from 1983-1997.
- After 1997, NA and JP sales spread out more, continuing to gradually increase and decrease.
- **EU sales consistently increase over time.**
- In the past 10 years (2006-2016), we are seeing sales change again - EU and JP sales are increasing, while NA sales are decreasing.

VIDEO GAME POPULARITY DATA PROJECT

RECOMMENDATIONS:

EU Sales - gradually increasing year over year

- Recommendation: Increase marketing spend on games/genres/publishers performing the best.

JP Sales - NA and JP sales have a negative correlation

- We can predict that what's popular in NA will not be popular in JP and what's not popular in NA will be popular in JP. I would dig further into the data to investigate and verify what genre, games, and publishers are most popular in JP. Once those values are established, I would recommend properly allocating the marketing budget to those items and increasing the marketing budget, as there may be room for more growth, especially based on the increase in sales in 2016.

NA Sales - have been gradually decreasing since 1997

- Recommendation: Is there additional room for growth in this region, or has this region reached its growth potential?

Links:

Project Reflections

View the full presentation

PREPARING FOR INFLUENZA SEASON

Background:

The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.

Objective:

Determine when to send staff, and how many, to each state.

Tools:

- Excel
- Tableau

Hypothesis:

If there is a higher percentage of the vulnerable population in certain states, then those states' hospitals would require more staff.

Data Sets:

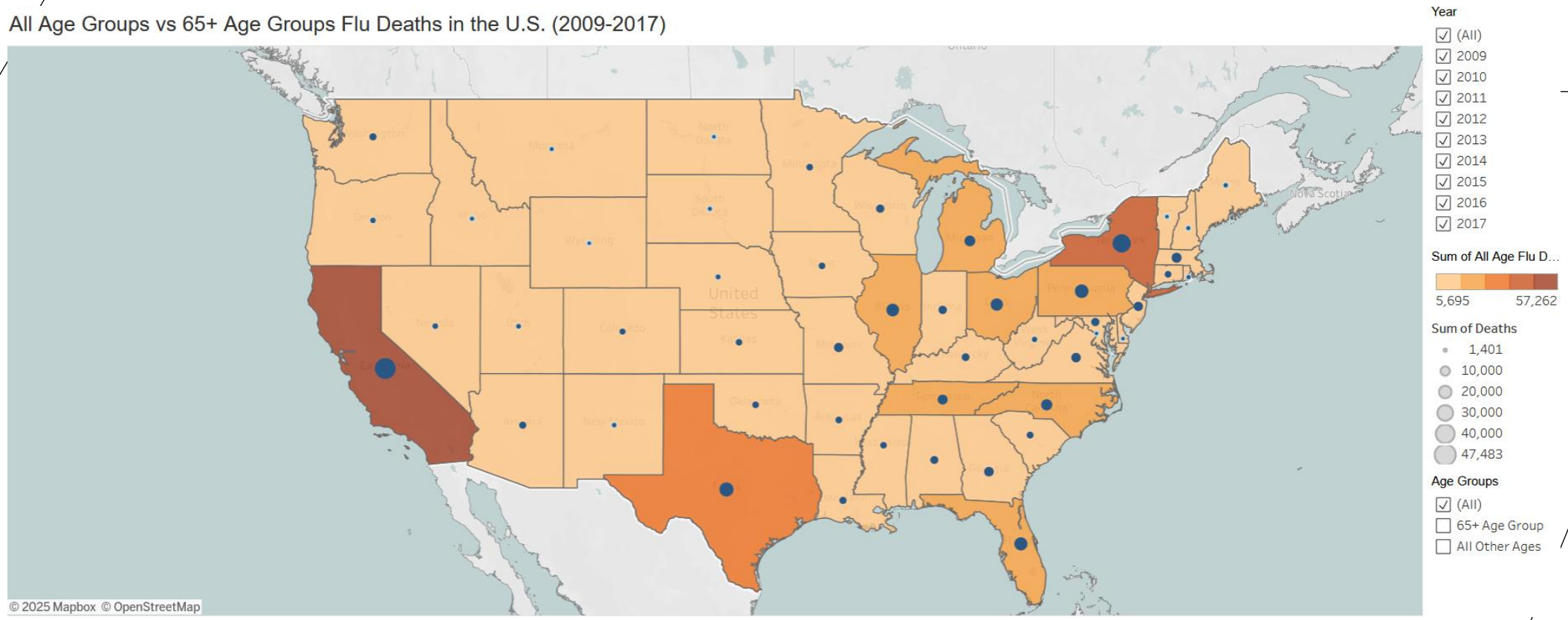
Influenza deaths by geography

- Data set source: CDC
- Summary of contents: This data shows the number of influenza deaths by ten-year age groups broken out by state and year.

Population data by geography, time, age, and gender

- Data set source: US Census Bureau
- Summary of contents: This data shows the estimated population totals for male, female, and 4-year age groups broken out by year and county/state.

PREPARING FOR INFLUENZA SEASON

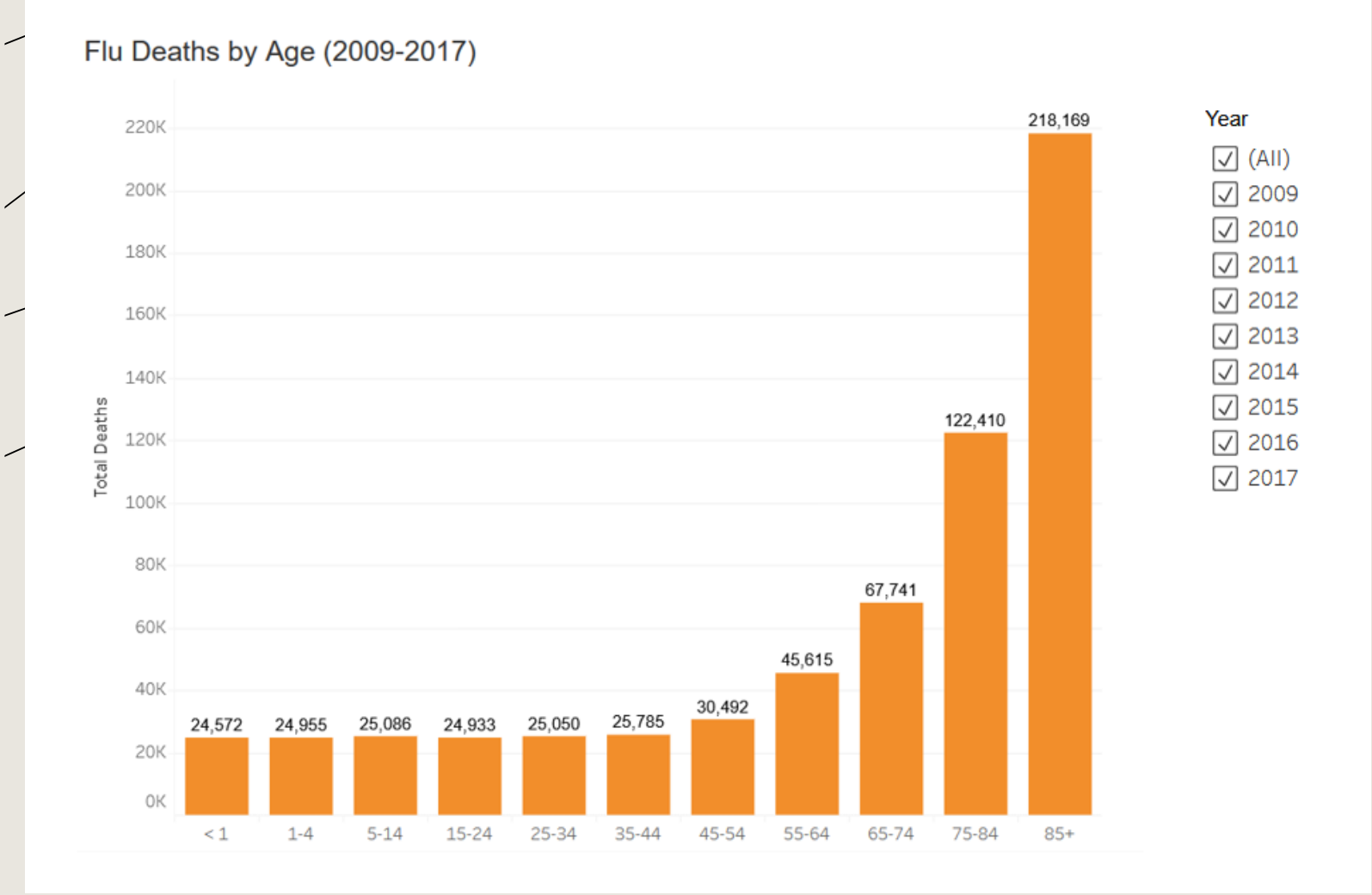


Combination map of all age groups flu deaths in the U.S. from 2009-2017

This map shows the total influenza deaths across the contiguous United States from 2009 – 2017.

The states with the highest deaths are **California**, **New York**, and **Texas**.

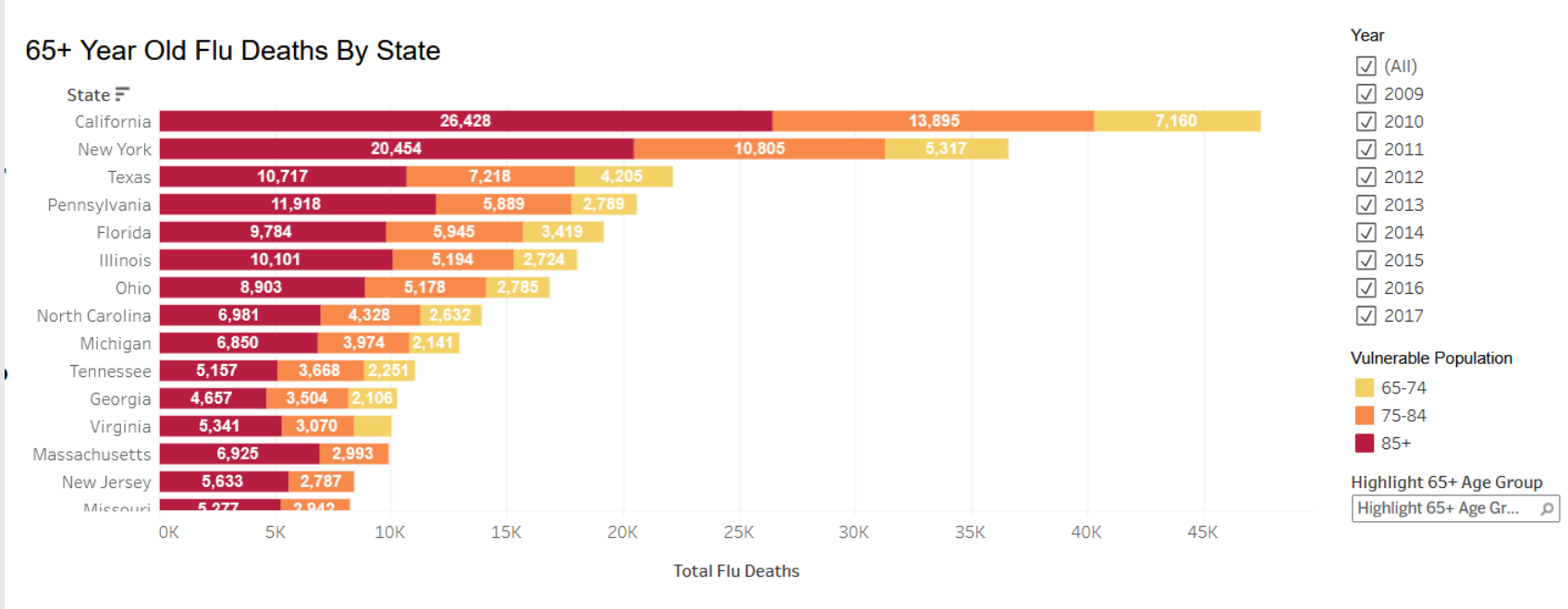
PREPARING FOR INFLUENZA SEASON



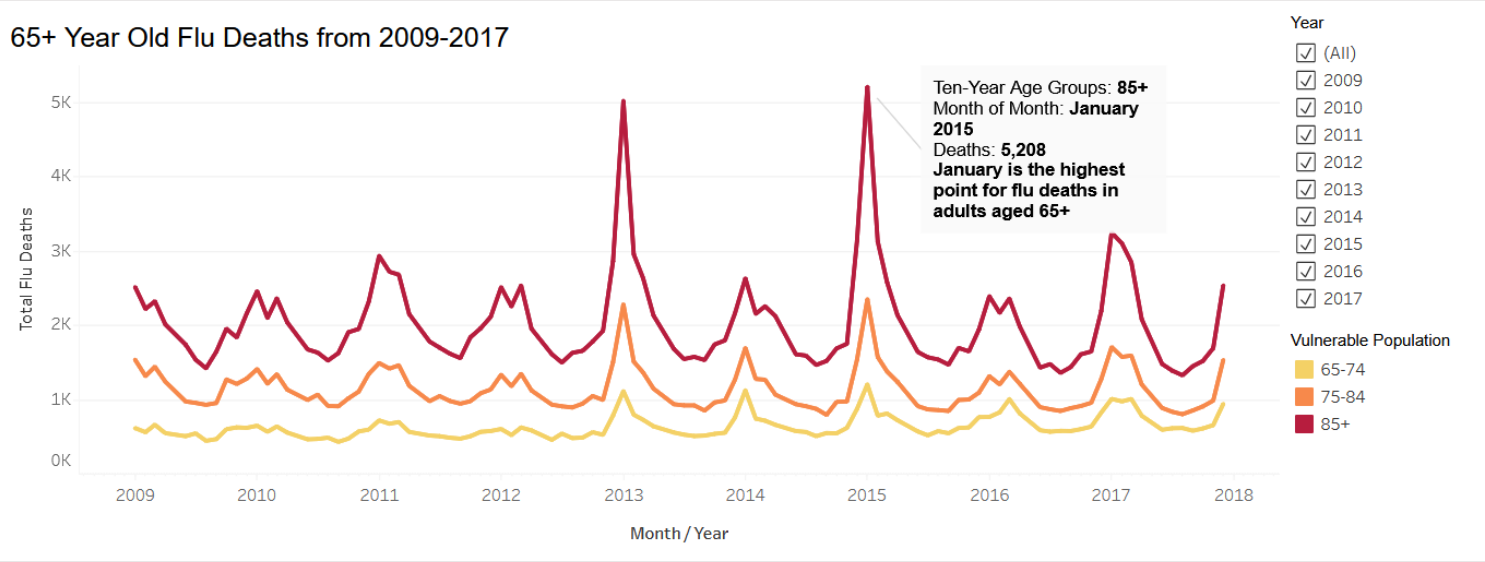
Bar chart of all age groups flu deaths in the U.S. from 2009-2017

- The data reveals that the most **at-risk groups are adults aged 65-74, 75-84, and 85 and older**.
- It is critical to prioritize sending additional healthcare workers to states with larger senior populations.

PREPARING FOR INFLUENZA SEASON



- The analysis shows that **California**, **New York**, and **Texas** has the highest populations of adults aged 65 and older who died of the flu. They also **rank among the top 5 most populated states overall** – further emphasizing the need for strategic staffing support.



- Influenza often peaks between December and February – with **January as the highest point** – flu-related deaths begin increasing as early as September/October and taper off by June/July.

PREPARING FOR INFLUENZA SEASON

RECOMMENDATIONS:

Prioritize sending seasonal healthcare staff to **California**, **New York**, and **Texas** – especially during the **peak flu months of December through February**.

Continuously **monitor all populations throughout the flu season** to reallocate healthcare staff as needed if flu-related deaths begin to rise in other regions.

Links:

- Influenza Deaths in Vulnerable Populations: Interim Report
- Tableau Storyboard
- Final Presentation
- Project Data Limitations and Metrics Documentation
- Descriptive Analysis: Conducting Statistical Analyses
- Hypothesis Testing: Conducting Statistical Hypothesis Testing

ROCKBUSTER STEALTH DATA ANALYSIS PROJECT

Background:

Rockbuster Stealth is a movie rental company that used to have stores around the world. Now, they are planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

Objectives:

Provide Rockbuster with data driven analysis to compete with other streaming platforms like Netflix and Amazon.

Tools:

- SQL
- PostgreSQL
- Tableau

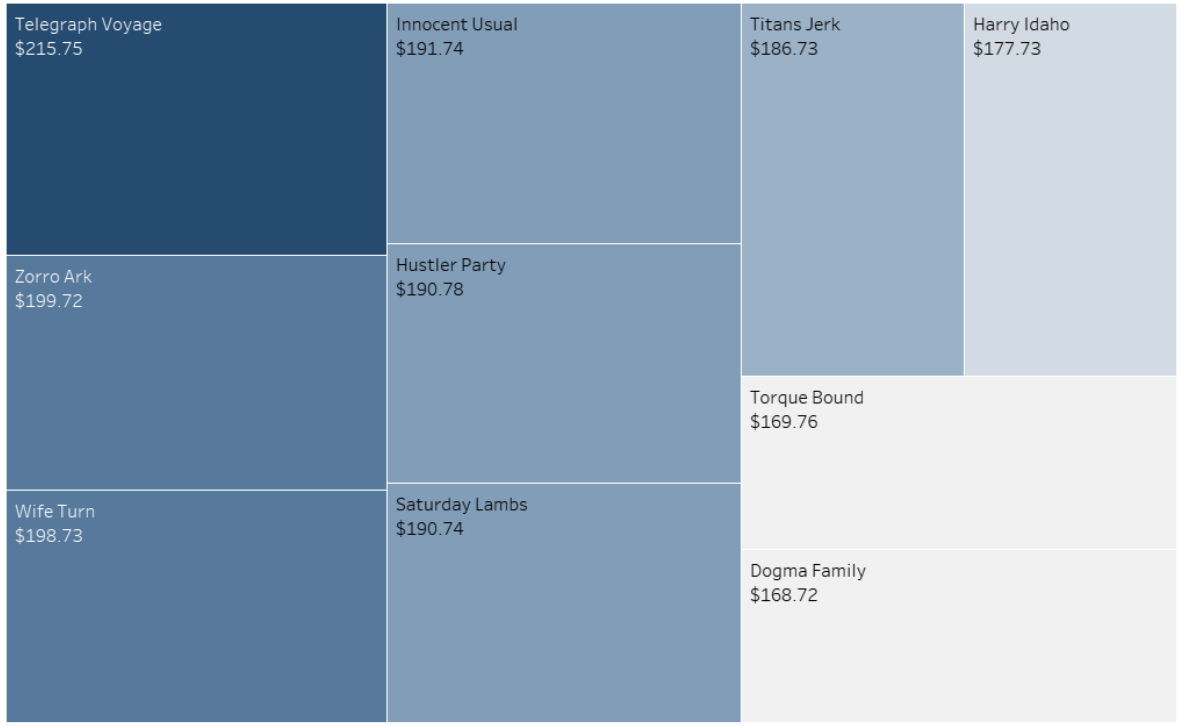
Key Questions:

- Which movies contributed the most/least to revenue gain?
- What was the average rental duration for all videos?
- Which countries are Rockbuster customers based in?
- Where are customers with a high lifetime value based?
- Do sales figures vary between geographic regions?

Data Set:

Rockbuster film inventory database

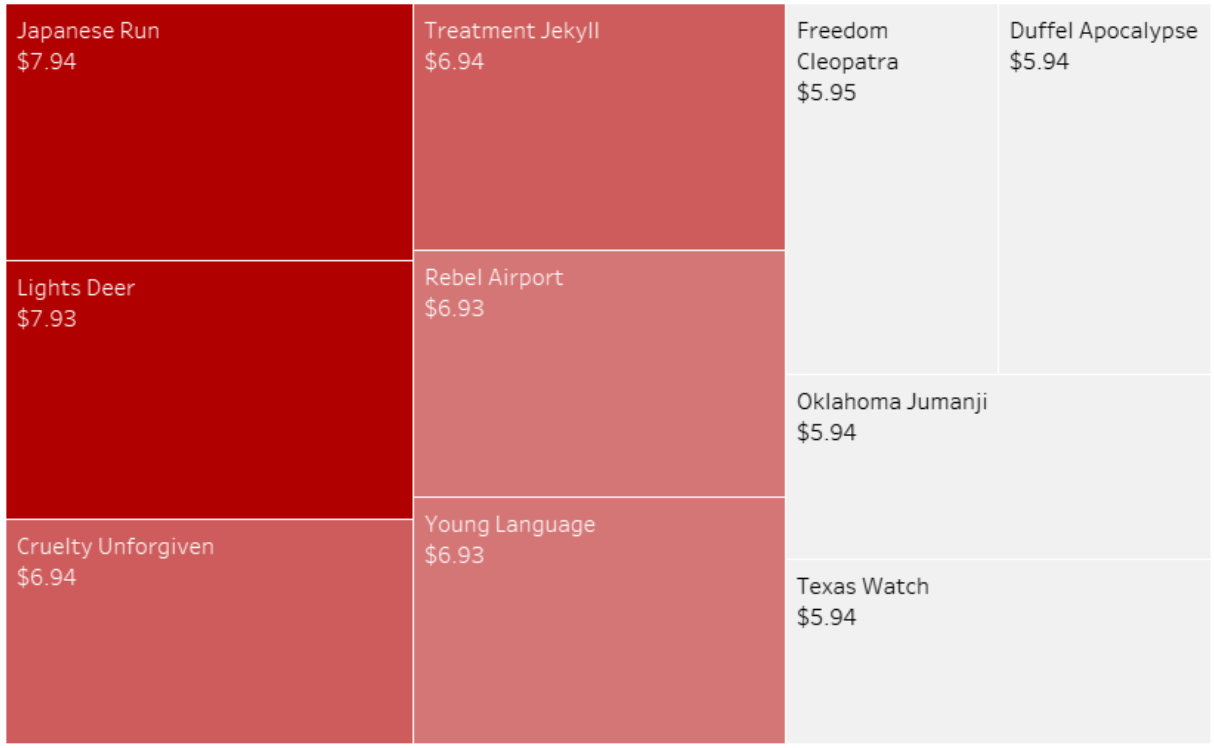
ROCKBUSTER STEALTH DATA ANALYSIS PROJECT



Tree map of highest grossing movies

These are the top 10 movies that contributed the most revenue, with **Telegraph Voyage** being the highest earner: $215.75.

These are the bottom 10 movies, which contributed the least revenue. The bottom three movies all made $5.94.



Tree map of lowest grossing movies

ROCKBUSTER STEALTH DATA ANALYSIS PROJECT

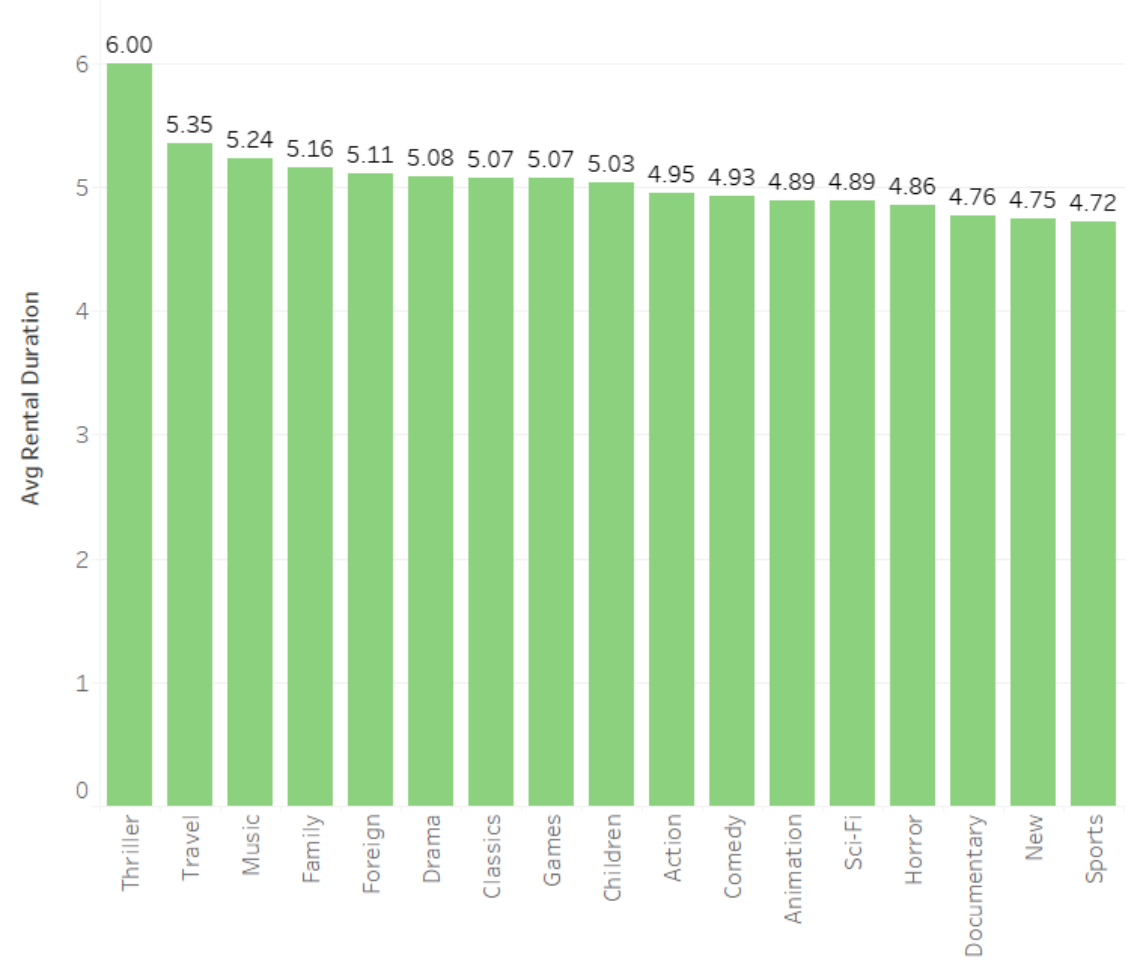
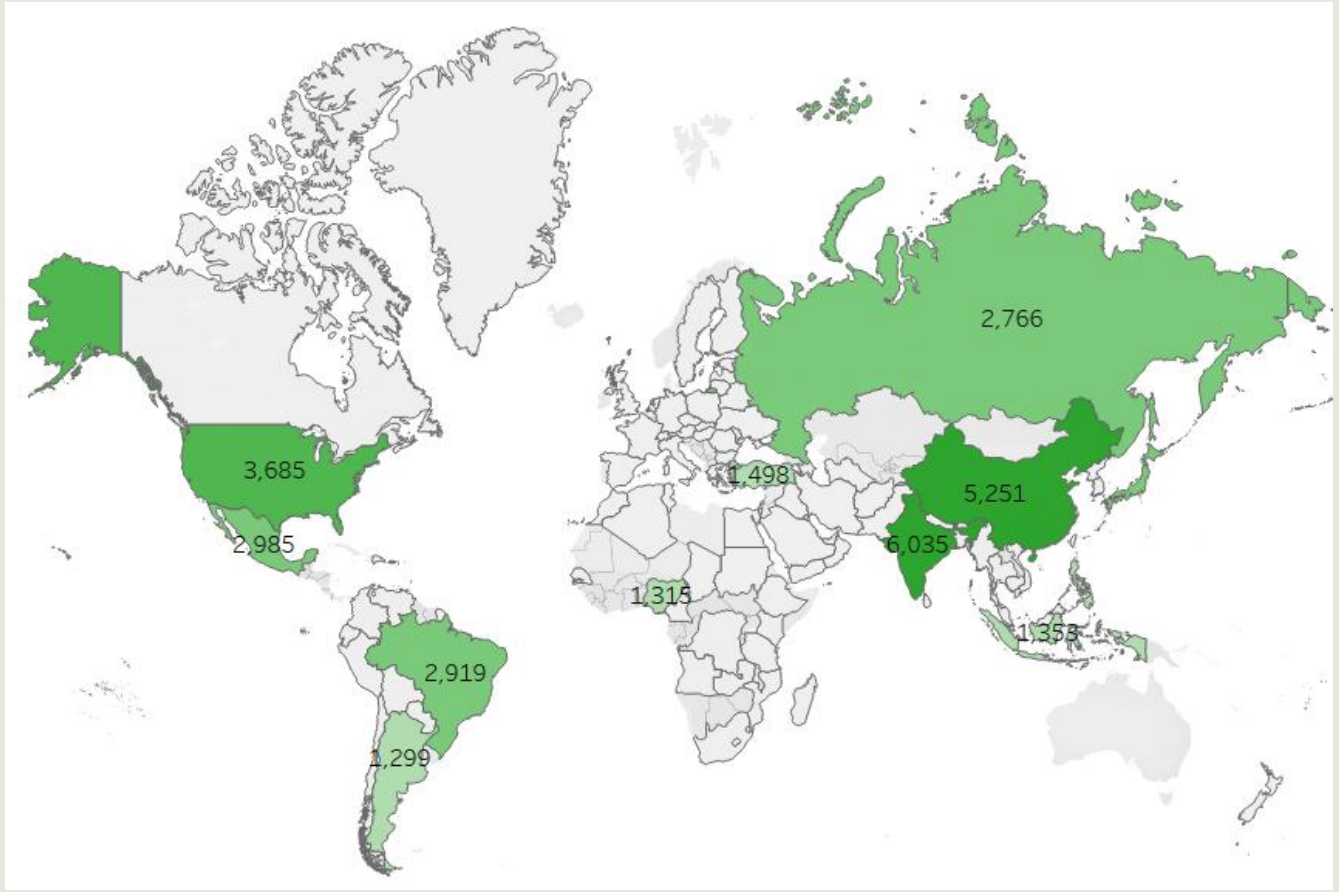


Chart showing average rental duration for each genre

The average rental duration for all movies was **4.985 days**.



Map showing how sales figures vary between geographic regions

Sales figures vary between geographic regions.

- **India** has the highest sales figures at $6,035.
- Followed by **China** with $5,251 and the **United States** with $3,685.

ROCKBUSTER STEALTH DATA ANALYSIS PROJECT

RECOMMENDATIONS:

- **Gain customer feedback via customer survey** – why choose Rockbuster over Netflix or Amazon? How to improve moving forward.
- **Track what users are watching** – this will give insight into what continues to be popular. Provide recommendations based on watch history.
- **Expand reach into different regions** – develop marketing budget and strategy.
- **Expand rental library** – have new releases available and expand already existing genre and rental library.

	Minimum	Maximum	Average	Count
Rental Price	0.99	4.99	2.98	1000
Rental Duration	3	7	4.985	1000
Movie Length	46	185	115.272	1000
Cost	9.99	29.99	19.984	1000
Most Frequent Rating		PG-13		

Chart of averages

Links:

Data Dictionary

Presentation

Technical Spreadsheet

Tableau Storyboard

Customer ID	First Name	Last Name	City	Country	Total Paid Amount
84	Sara	Perry	Atlixco	Mexico	$128.70
518	Gabriel	Harder	Sivas	Turkey	$108.75
587	Sergio	Stanfield	Celaya	Mexico	$102.76
537	Clinton	Buford	Aurora	United States	$98.76
367	Adam	Gooch	Adoni	India	$97.80

Chart of customers with high lifetime value and where they are located

INSTACART GROCERY BASKET ANALYSIS

Background:

Instacart is an online grocery store that operates through an app. While Instacart already has good sales, they are look to uncover more information about their sales patterns.

Objective:

Perform an initial data and exploratory analysis of some of Instacart's data in order to derive insights and suggest strategies for better segmentation.

Tools:

- Python
- Excel

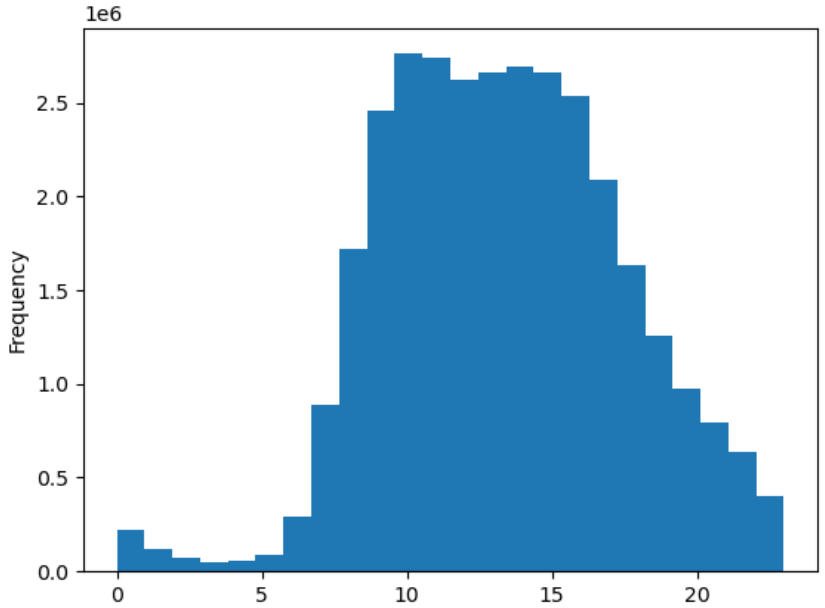
Key Questions:

- The sales team needs to know what the busiest days of the week and hours of the day are (i.e., the days and times with the most orders).
- They also want to know whether there are particular times of the day when people spend the most money.
- What's the distribution among users in regards to their brand loyalty (i.e., how often do they return to Instacart)?
- Is there a connection between age and family status in terms of ordering habits?

Data Set:

- Open-source data sets from Instacart
- Customer data set created by CareerFoundry

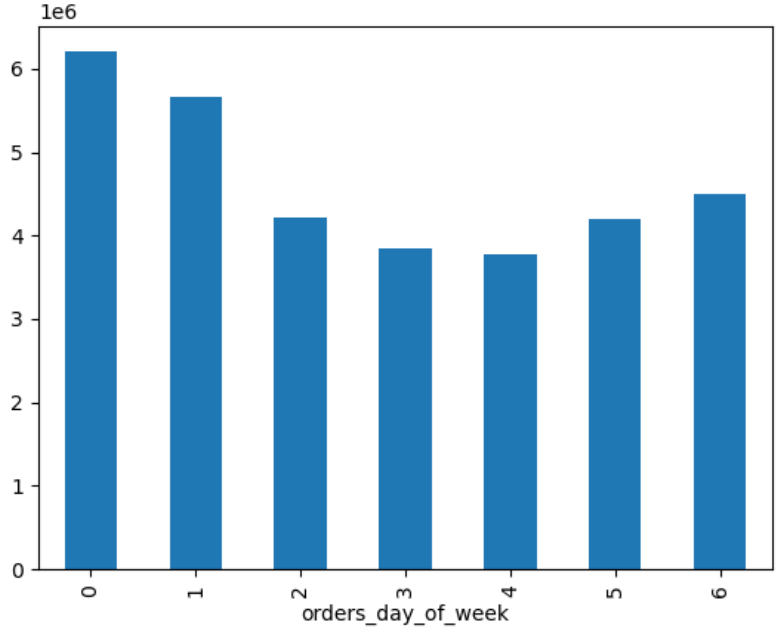
INSTACART GROCERY BASKET ANALYSIS



Histogram showing orders during a 24-hour day

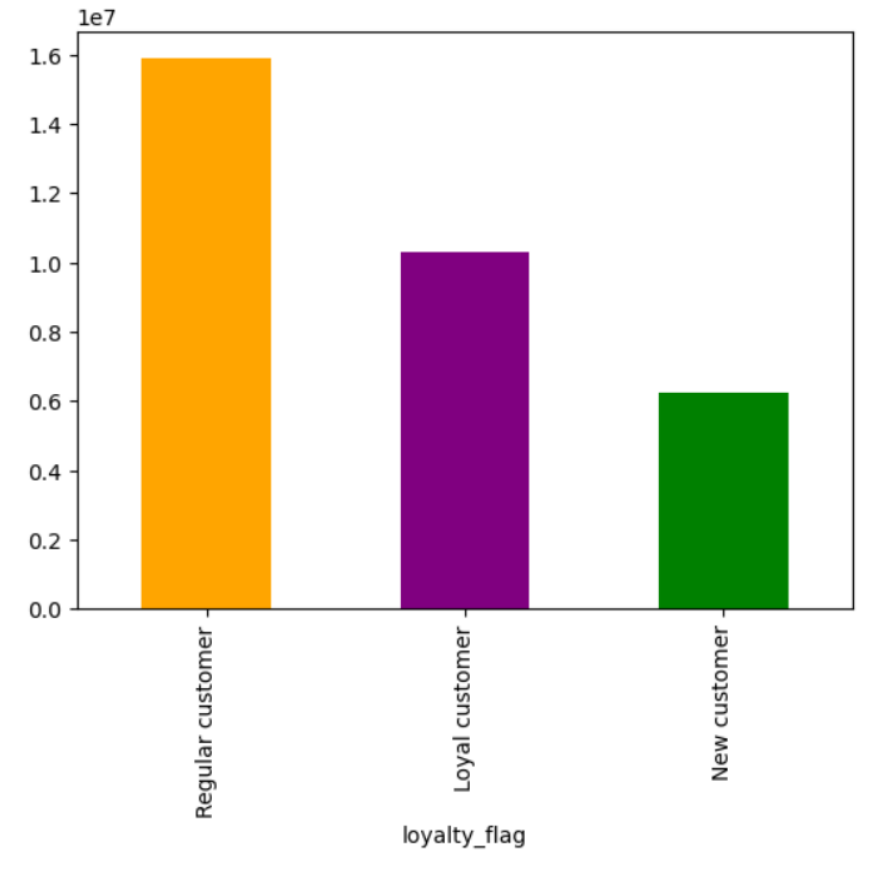
Most orders come through during the **11:00 to 15:00 hours** of the day. Instacart can expect to be busiest during those timeframes and should aim to schedule ads during the hours of 0:00 - 10:00 and 16:00 - 24:00.

This bar chart shows the orders during the day of the week. 0 = Saturday and 1 = Sunday, so **weekend shopping is the most popular**.



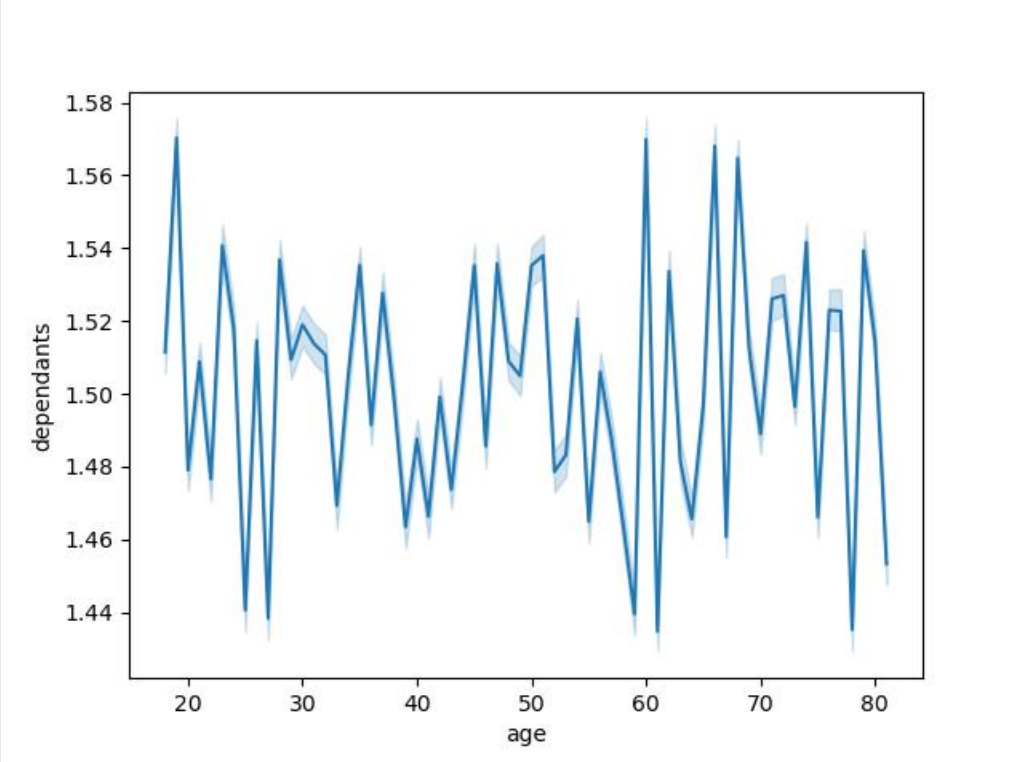
Bar chart showing orders during the week

INSTACART GROCERY BASKET ANALYSIS



Bar chart showing loyalty flag

Regular customers are the primary customer at Instacart. Their max orders are less than or equal to 40 and greater than 10. Loyal customer are the second highest customer population, with max orders over 40. New customers also have a large chunk of the customer database, with less than or equal to 10 max orders. The new customer segment could be an excellent growth opportunity.



Line chart showing correlation between dependents and age

There is **no connection** between dependents and age, as there is no overall direction or consistency in the line chart. There is no discernible pattern within the chart.

INSTACART GROCERY BASKET ANALYSIS

RECOMMENDATIONS:

- It is recommended to serve ads during the 1:00 - 10:00 hours and the 16:00 - 24:00 hours, with the most ads 1-2 hours before the prime shopping window of 9:00 - 10:00.
- As the price distribution is primarily steady throughout the day, my recommendation is to follow the order hour of the day and prioritize advertising high value products during the 9:00 and 10:00 window.
- Is there any loyalty perks that could be gained if a customer converts from a regular customer to a loyal customer? If not, it may be beneficial to create one to gain additional loyal customers.
- As there is no connection between dependents and age, there is no recommendation. Perhaps understanding a different relationship would yield better results.

Link:

View final report

LIBRARY FUNDING AND LITERACY RATES

Background:

For this project, the goal was to determine if public libraries and public library funding help close the literacy gap.

Objectives:

Provide data driven analysis to determine if public libraries are literacy drivers.

Tools:

- Python
- Tableau

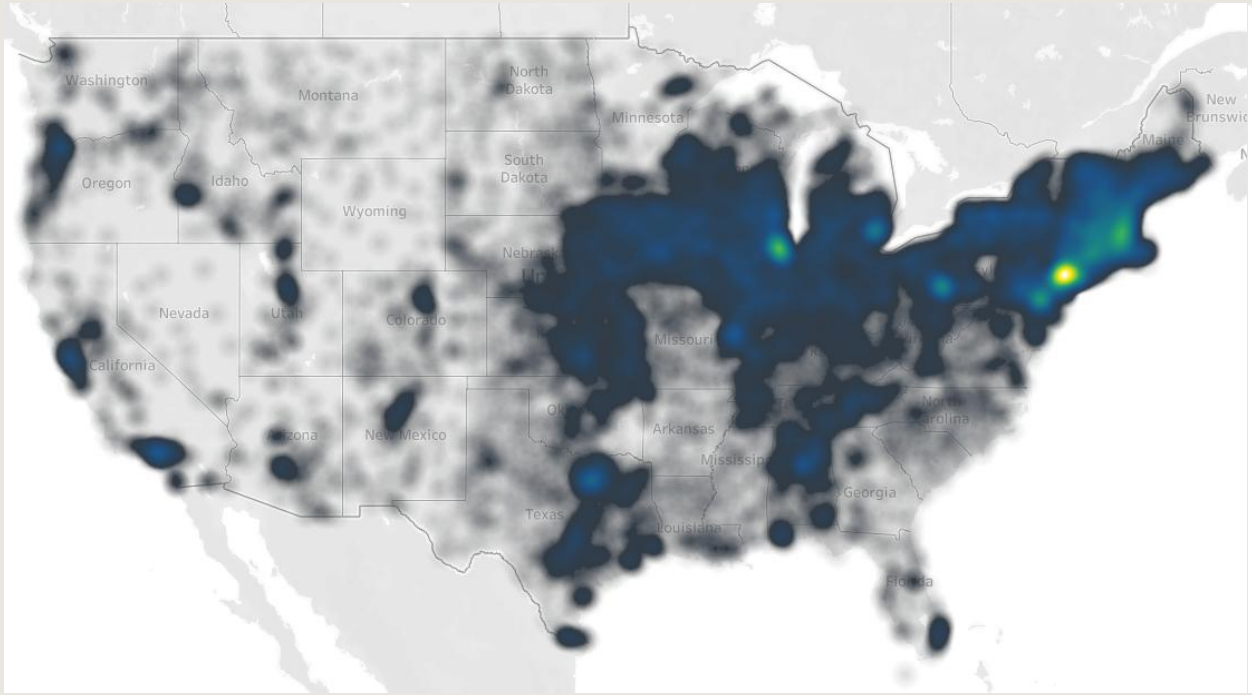
Data Sets:

- Public Libraries Survey (PLS)
- U.S. Skills Map: State and County Indicators of Adult Literacy and Numeracy

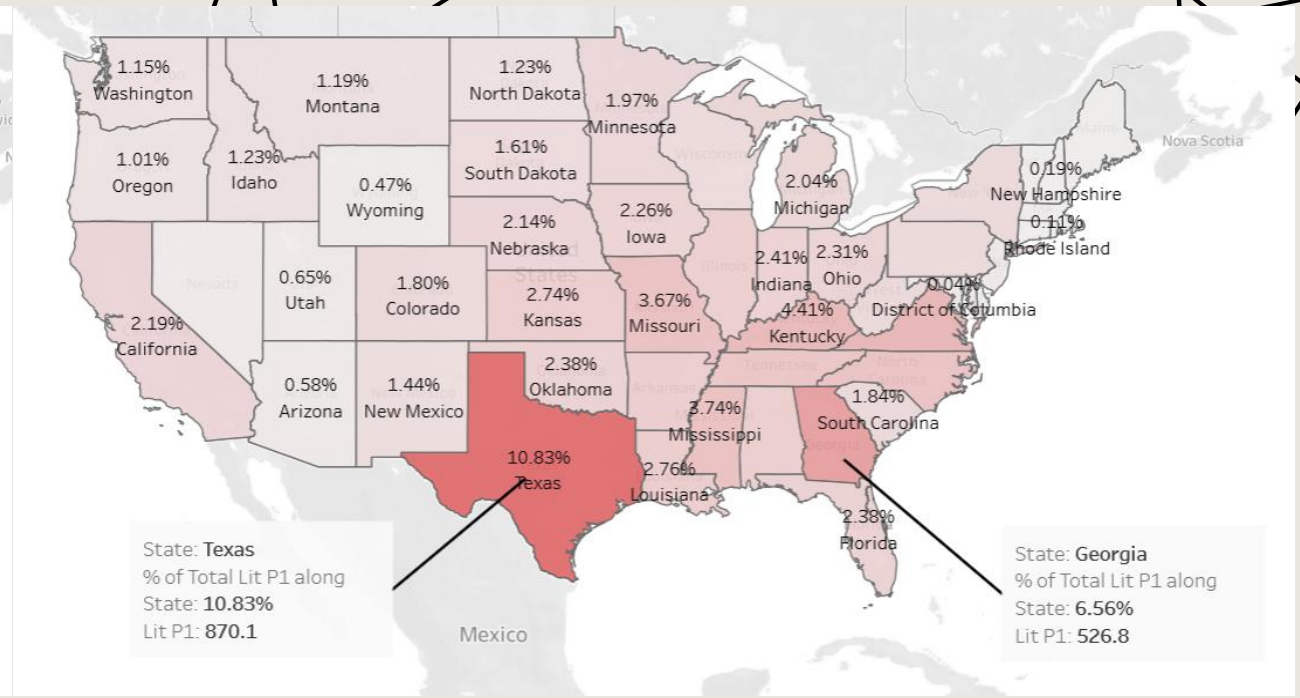
Key Questions:

- Is there a relationship between library funding and literacy rates across the U.S. counties?
- Do communities with higher program attendance see higher literacy rates?
- Do libraries with a higher check-out rate see higher literacy rates?
- How does income level correlate with literacy rates across communities?
- Are there low-income counties with unexpectedly high literacy rates – and do library resources play a role?

LIBRARY FUNDING AND LITERACY RATES



Map showing density of public library locations in U.S.



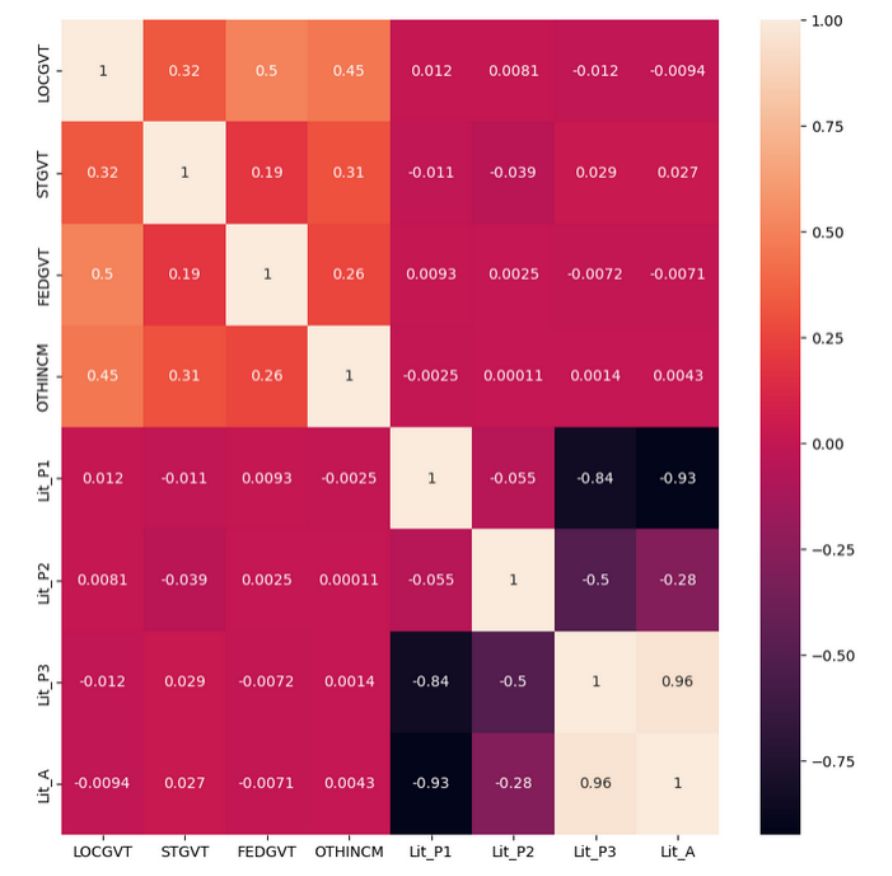
Map showing percentage of adults (aged 16-74) in U.S. with low literacy

For this project, the goal was to determine if public libraries and public library funding helps close the literacy gap.

The density map shows the density of public libraries in the U.S. The **most densely populated regions are the North and Northeast**.

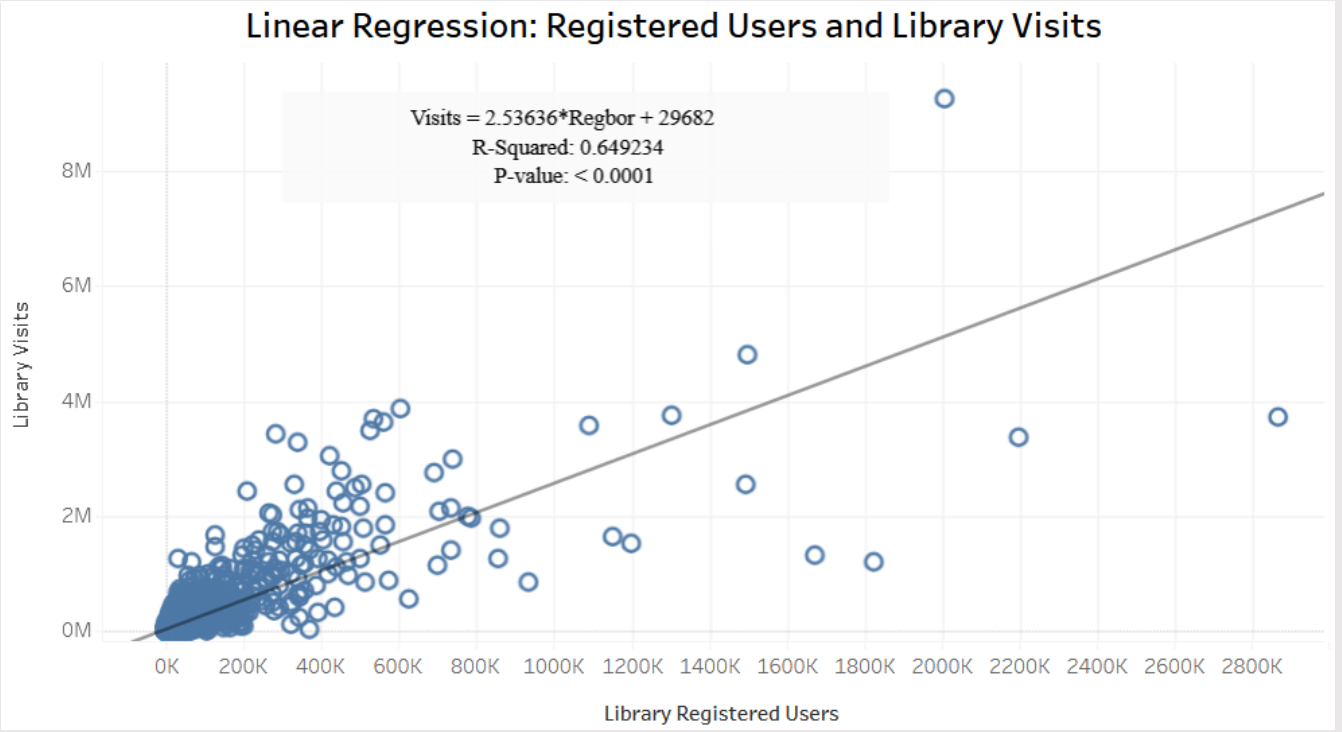
The other map shows adults (16-74) with low literacy. **Texas is the state with the highest percentage of adults with low literacy (10.83%), followed by Georgia (6.56%).**

LIBRARY FUNDING AND LITERACY RATES



Correlation heatmap

This correlation heatmap is correlating government funding and literacy rates. All correlations are basically at zero, meaning that there is **no correlation between any level of government funding and literacy rates**.



Linear regression chart

While conducting linear regression between library visits and library registered users, you can see the logical relationship – the more registered users a library has, the more visits. While there is a relationship between these two values, **there is still a lot of information left to uncover**.

LIBRARY FUNDING AND LITERACY RATES

TO NOTE:

The county data set used in this analysis had a large amount of data missing, so imputation was used to fill in the missing values. This made the data set incomplete and did not yield significant results.

NEXT STEPS:

- **Find a better data set** that has more complete data on county demographics.
- While some questions have been partially answered, **it's crucial to continue analysis** and focus on answering the research questions.

Links:

Tableau Storyboard

GitHub

THANK YOU

Rebecca Meyer

rlwinterisu@gmail.com

GitHub

Tableau