

# Data Storytelling

Dr. Patrick Chester

Fall 2026

Week 02

# Announcements

# Announcements

- Who here has used GitHub before?
  - If there are no major objections, slides and other course materials will be hosted on GitHub this semester.
- If the room is too small, let me know.. I'll discuss moving to a larger room with the department.

# Data Storytelling

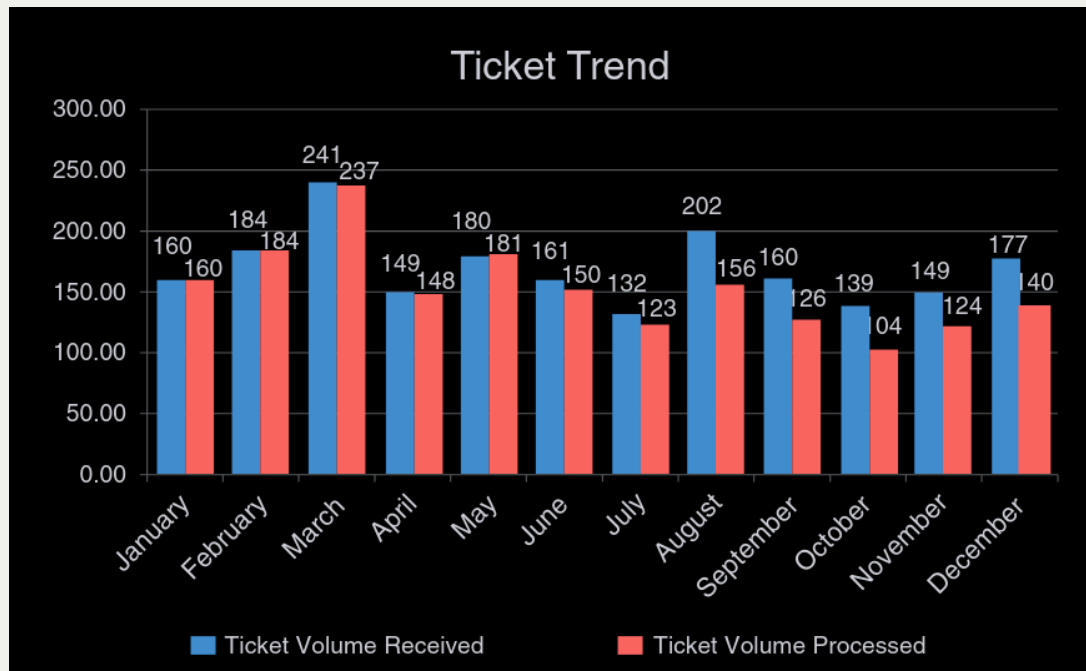
# Why tell stories with data?

- This class represents a wide swath of the Stevens campus and its diverse academic disciplines:
  - Engineers, Social Scientists, Natural Scientists, Business Majors, Humanities Majors, and More!
- No matter your discipline, you will need to communicate your ideas, theories, models, and findings with others.
- Data storytelling is effectively the art of communicating data-driven insights in a compelling and understandable way.

# Challenges of Data Communication

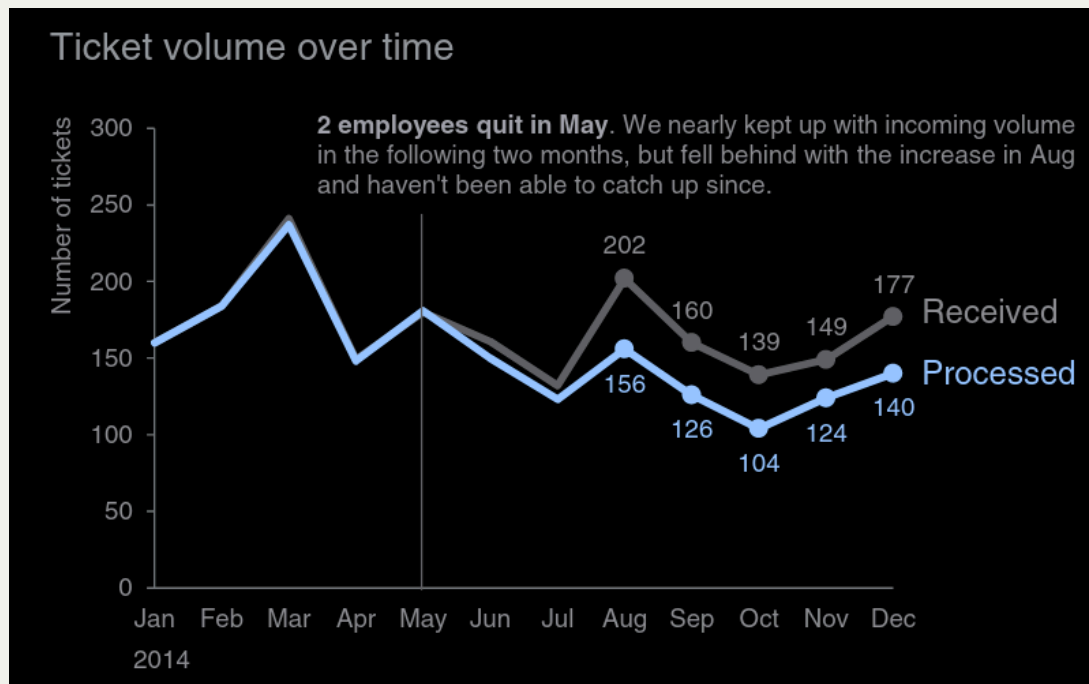
- While data can provide powerful insights, communicating these insights effectively can be challenging.
- It is still common to see poorly designed charts, misleading graphs, and data presented without context.

# Illustrative Example 1: Confusing Graph



- What is this graph trying to communicate?
- How could it be improved?

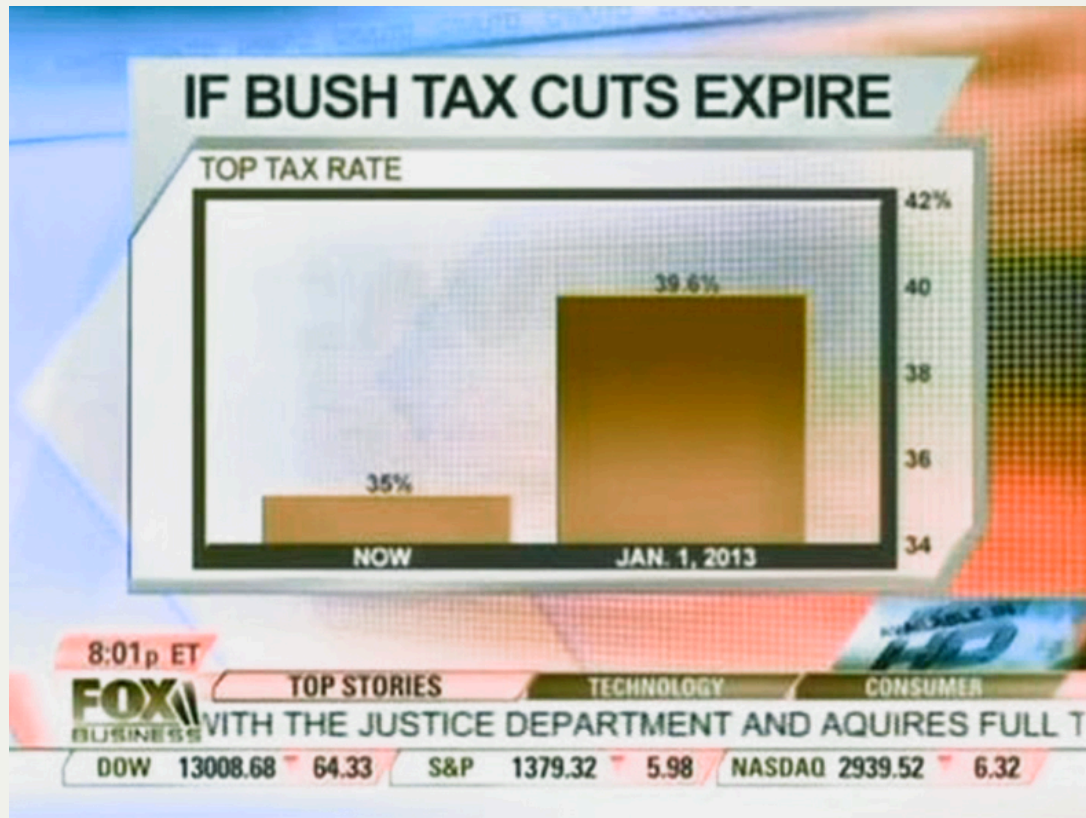
# Illustrative Example 1: Improved Version



- Is this graphs informational content equivalent to the previous graph?
- Does this communicate the information more clearly? Efficiently?

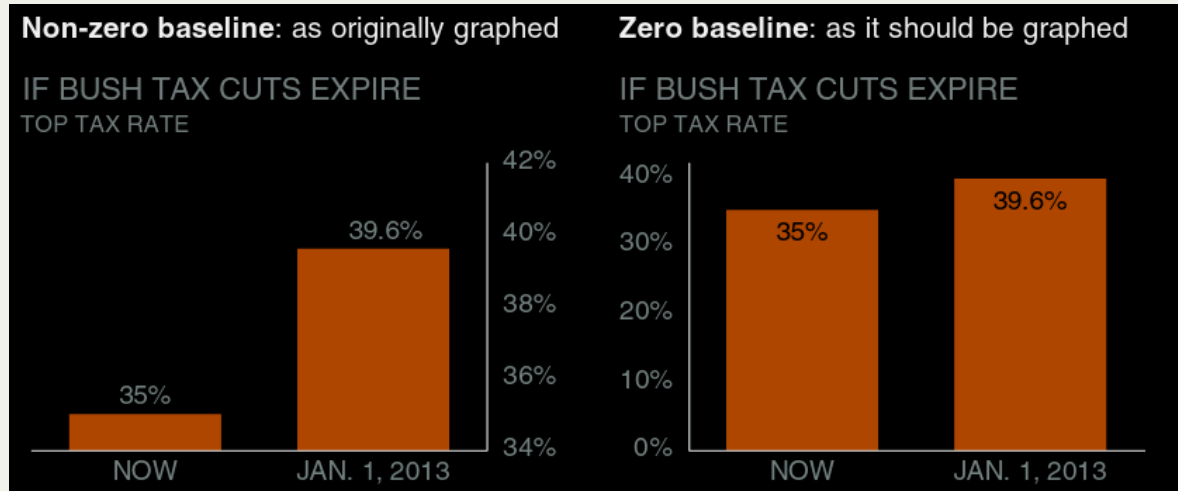


# Illustrative Example 2: Misleading Graph



- What is this graph trying to communicate?
- How could it be improved?

# Illustrative Example 2: Improved Version



- Is this graphs informational content equivalent to the previous graph?
- Is this a more honest representation of the data?

# Objectives of This Course

- Equip you with the technical skills to analyze data using R.
- Teach you to avoid pitfalls and follow ethics of data analysis.
- Prepare you to collaborate with others on data-driven projects.
- Teach you how to create compelling data visualizations.
- Foster critical thinking about data and its presentation.

# Objectives of Analysis

# What are we doing here, anyway?

- Before starting any project, you must first start with a clear objective
  - Do you hope to gain a better understanding of some phenomenon?
  - Do you want to test a hypothesis?

# Exploratory Analysis

- Also known as descriptive analysis.
- Investigating data to find patterns, trends, or relationships without a specific hypothesis in mind.
- *Commonly used by:*
  - Scientists looking to gain a better understanding of the dynamics of datasets.
  - Engineers trying to understand the performance of a model or system.
  - Journalists investigating data for stories.

# Exploratory Analysis Process

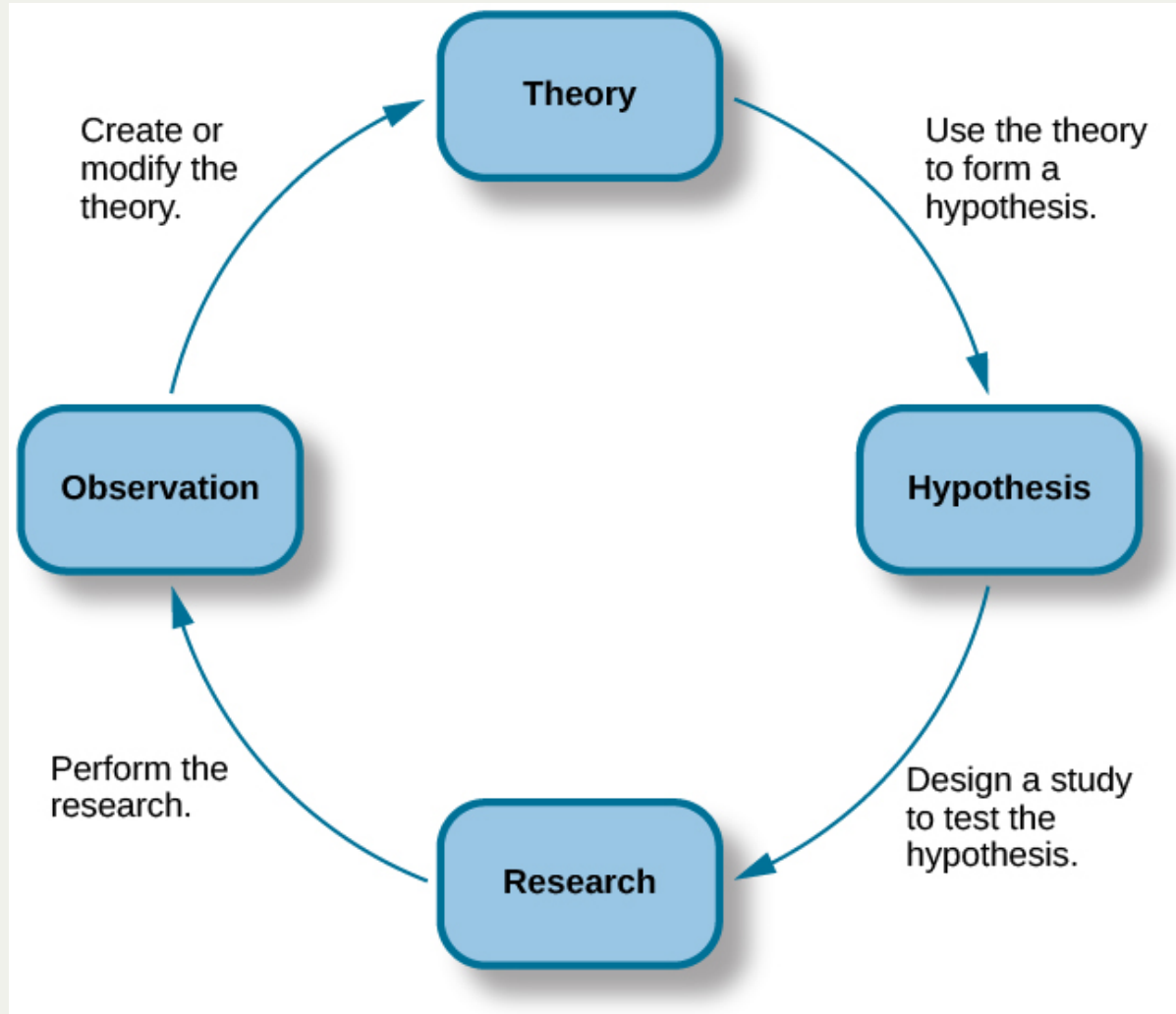
1. *Data Collection*: Gather relevant data from various sources.
2. *Data Cleaning*: Handle missing values, outliers, and inconsistencies.
3. *Data Visualization*: Create charts and graphs to visualize data distributions and relationships.
4. *Pattern Recognition*: Identify trends, correlations, and anomalies in the data.
5. *Hypothesis Generation*: Formulate hypotheses based on observed patterns for further testing.

# Explanatory Analysis

- Also known as confirmatory analysis, this involves testing a specific hypothesis or theory using data.
- *Commonly used by:*
  - Scientists testing theories or models.
  - Business analysts evaluating the effectiveness of strategies.
  - Engineers validating whether a system meets design specifications.



# Explanatory Analysis Process



# The Who, What, and How of Data Storytelling

# The Fundamentals of Data Storytelling

- To effectively communicate data-driven insights, consider the following three key questions:
  1. Who is your audience?
  2. What do you want your audience to know or do?
  3. How can you use data to deliver that message?

# The Who: Audience Analysis

- Who is it to whom you're telling the story?
  - Professionals in your field?
  - General public?
  - Decision-makers?
- Understanding your audience's background, interests, and understanding of the subject is crucial for tailoring your message effectively.

# The Who: Audience Analysis

- Examples of audience considerations:
  - Technical vs. Non-Technical: Adjust the complexity of your language and explanations.
    - Ex. Academic audiences may appreciate technical jargon, while general audiences may need simpler explanations.
  - Interests: Focus on aspects of the data that are most relevant to your audience's concerns.
    - Ex. Academics will care about methodology and causation, while business leaders may care more about implications and actions.
  - Prior Knowledge: Gauge how much background information is needed to understand your message.
    - Ex. An expert audience may not need basic concepts explained, while novices may require more context.

# The Who: Your Interests

- What are you trying to achieve with your data story?
  - Are you trying to inform, persuade, or motivate action?
- Clarifying your own objectives will help you focus your message and choose the most relevant data to include.

# The What: Crafting Your Message

- What do you want your audience to learn or do after hearing your data story?
  - Identify the key insights or takeaways you want to convey.
- You likely *are* the foremost expert on your topic in the room!
  - Use your expertise to highlight the most important aspects of the data to those who may not be as familiar with it.

# The What: Crafting Your Message

- How will you be communicating your message?
  - Will you be using visualizations, narratives, or a combination of both?
- Choose the format that best suits your audience and objectives.
  - Graphs are helpful for communicating patterns and trends.
    - Ex. A time series graph of GDP over time.
  - Tables are useful if you want to share organized information for reference.
    - Ex. A table of summary statistics or regression coefficients.



# Presenting Data Effectively

- Who here has ever given a presentation to an audience with slides?
- *Dos*
  - Use clear, concise titles and labels.
  - **Highlight** key points with **color** or **emphasis**.
  - Practice your delivery to ensure clarity and confidence.
- *Don'ts*
  - Avoid cluttered slides with too much text or too many graphics.

# The How: Tools for Data Storytelling

## Data

- Once you have a question/hypothesis, the next task is to gather data that will help you to either answer your question or test your hypothesis.
- Data can come from a variety of sources:
  - Public Datasets (e.g., government databases, research repositories)
  - Surveys and Questionnaires
  - Experiments
  - Web Scraping
  - APIs
- We will explore how to gather data in a subsequent class

# Examples

- Let's say that you want to better understand how life expectancy varies across different countries.
  - What data would you need to gather and what sources could be of value?
  - You might gather data from the World Bank or World Health Organization.
- Let's say you want to see how regime type (e.g., democracy vs. autocracy) affects economic growth.
  - What data would you need to gather and what sources could be of value?
  - You might gather data from the Polity Project or Freedom House for regime type and World Bank for economic data.

# Side Note: Data Ethics

- It is always tempting to only look for data that supports your hypothesis or desired outcome.
  - However, the goal of any data analysis should be to seek the truth, whatever that truth may be.
- Always be transparent about your data sources, methods, and any limitations of your analysis.
- Misrepresenting data, cherry-picking results, or engaging in fraud can lead to disastrous outcomes at the personal and societal levels.

# Side Note: Data Ethics

## Individual-level

- Those engaging in this action are likely to be eventually caught and face serious consequences.
  - Ex. Michael LaCour's fraudulent study on changing attitudes toward LGPTQ

## Societal-level

- Policy-makers and individuals may make poor decisions based on inaccurate or misleading data.
  - Ex. Untold numbers of people died of cigarette-related illnesses due to a tobacco industry cover-up of health risks.

# Data Storytelling: An Illustrative Example

- Let's say that you are an engineer working for a car manufacturer. You believe that a recent change in the design of your cars is increasing the rate of car accidents.
  - Who is your audience?
  - What do you want them to know or do?
  - How can you use data to deliver that message?

# Setting up your system for data analysis

# Preparation for Data Analysis with R and RStudio

- Next class, we will be starting to do some hands-on data analysis using R and RStudio.
- To do that, you will need to have both R and RStudio installed on your computer prior to the next class.
- Let's get that set up now!



# Installing R

- Go to <https://cran.r-project.org/mirrors.html>
  - **Window users:** Click on “Download R for Windows” then “base” then “Download R x.x.x for Windows”
  - **Mac users:** Click on “Download R for (Mac) OS X”
    - Can also install via Homebrew: `brew install --cask r`
  - **Linux users:** `sudo [apt/dnf/yum] install r-base`

# Installing RStudio

- Got to <https://posit.co/download/rstudio-desktop/>
  - Download the free version of RStudio Desktop for your operating system.
  - Follow the installation instructions for your OS.

# GitHub Desktop

- Another critical tool for data analysis and collaboration is GitHub.
- Go to <https://desktop.github.com/download/>
  - Download and install GitHub Desktop for your operating system.