

Class 1: Introduction to Business Forecasting

Krzysztof Zaremba

Who am I?

- **Krzysztof Zaremba**
 - **Education:** PhD in Economics, Columbia University
 - **Research Focus:** Applied Econometrics, Health Economics
 - **Email:** zaremba@itam.mx
 - **Office:** Santa Teresa Campus
 - **Office Hours:** Mondays 2–3pm on Zoom (email for other times)

What Is This Class About?

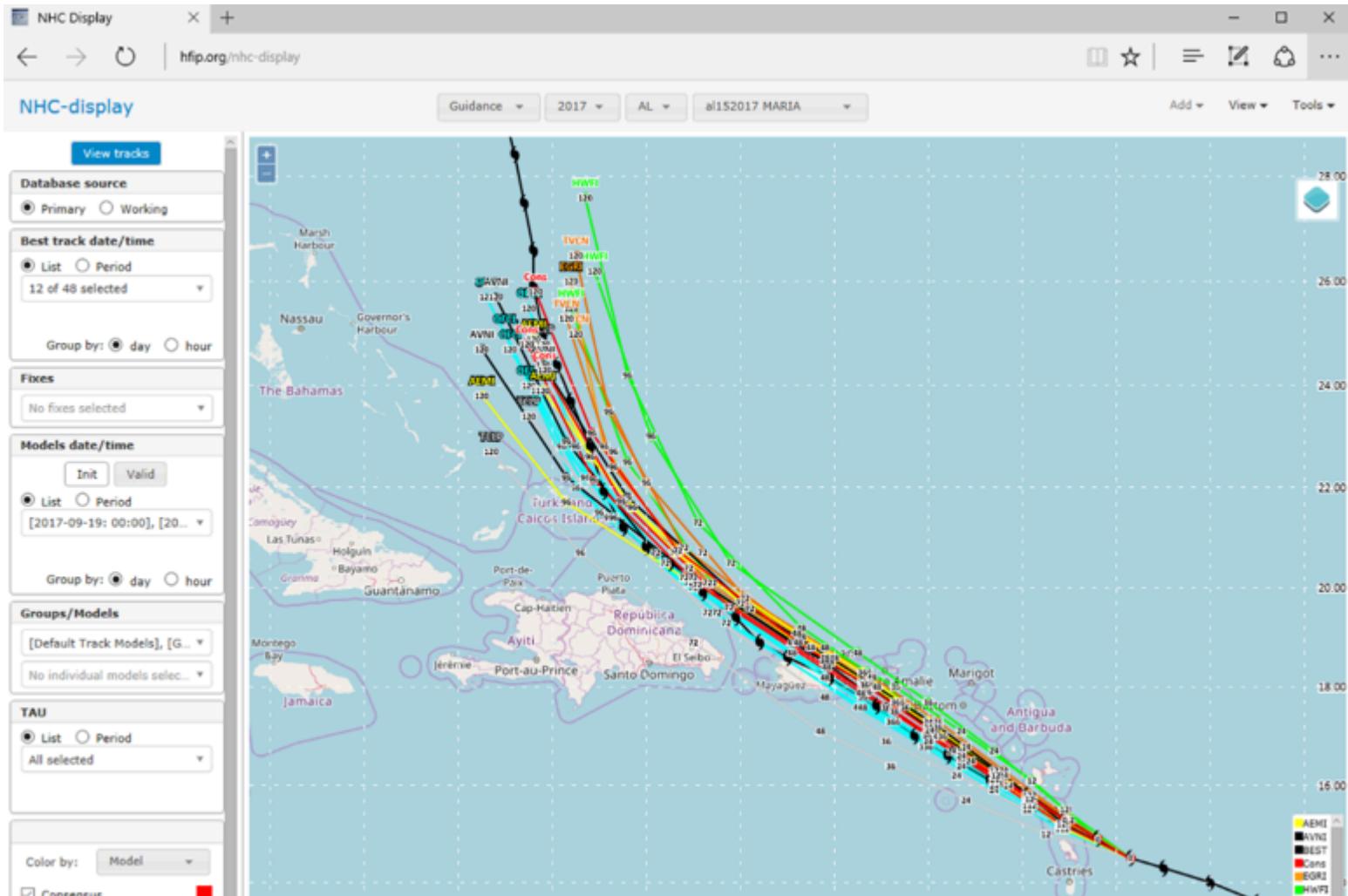
Forecasting is the science and art of making predictions about future events using historical data and relevant information.

In this class, you will learn how to turn data into actionable insights that drive better decisions in business and beyond.

Why is Forecasting Important?

1. Weather Forecasting

- Predicting hurricanes → timely reactions and preparations.



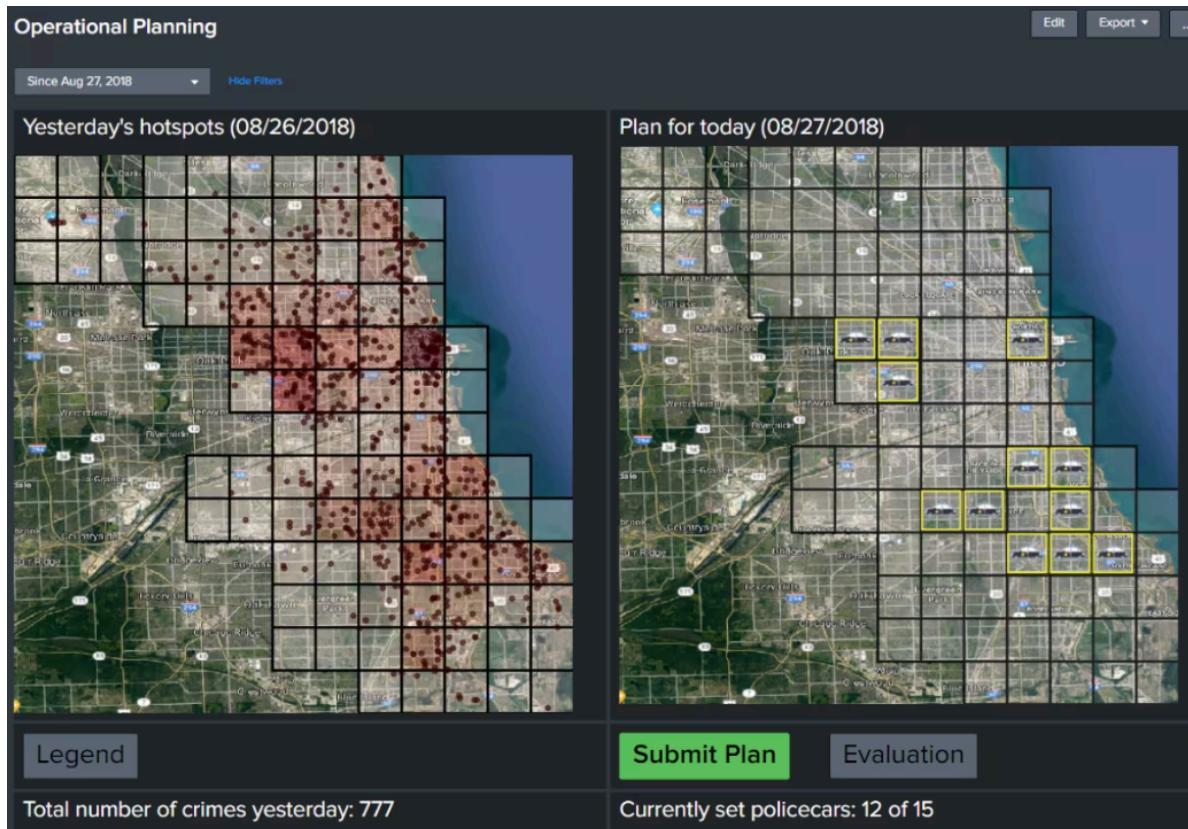
1. Epidemic Forecasting

- Predicting disease outbreaks → effective public health responses.

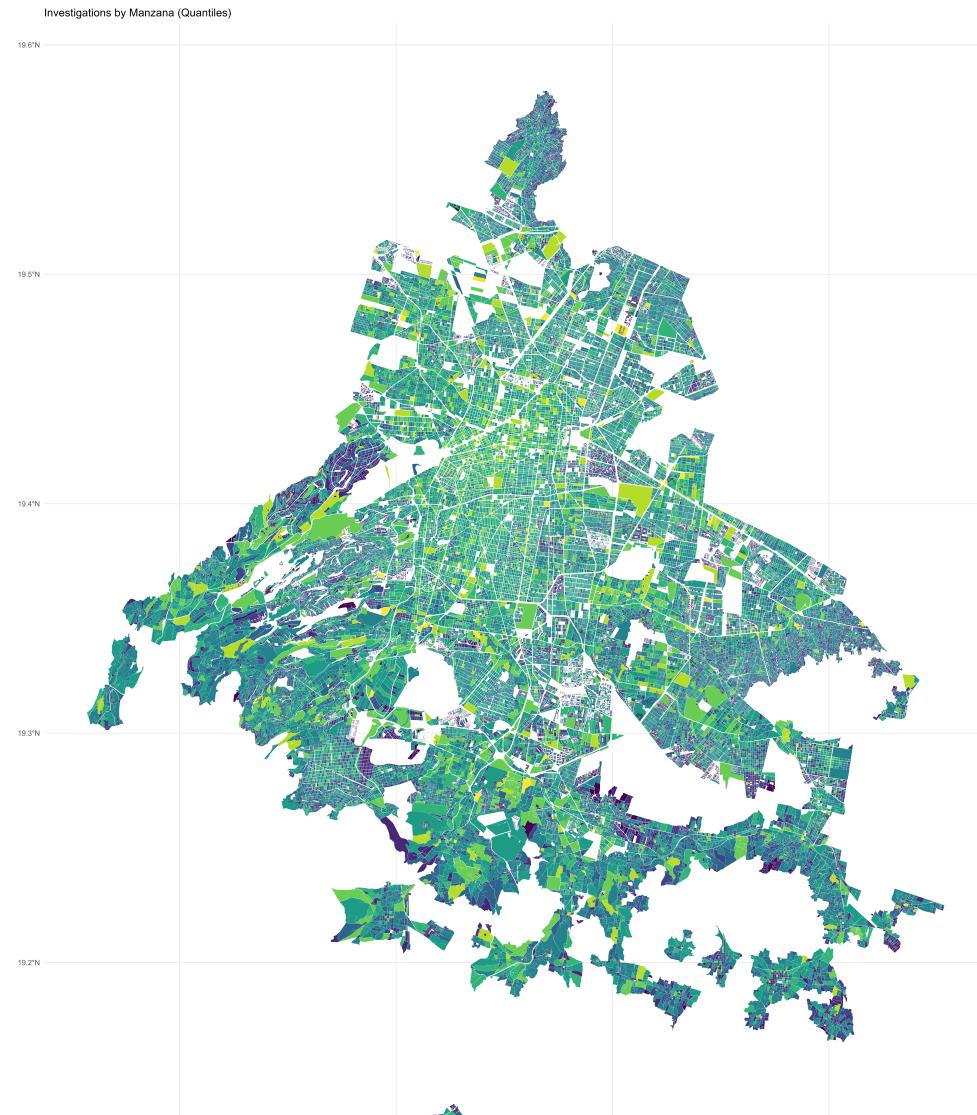


1. Crime Forecasting

- Identifying crime hot spots → more effective police patrols.
- Identifying individuals at risk → targeted interventions.



One of your tasks later in class: **Predict 911 calls in Mexico City**



What Is This Class About?

Forecasting in Business

- With the increasing availability of data, firms rely on rigorous data-driven methods to make decisions.
- Many business decisions depend on anticipating future events.
- In this course, you'll learn how to use data to make informed predictions about the future.

Applications of Forecasting

1. Anticipating Demand

Adjust inventory to minimize waste and meet customer needs effectively

Example: Meal kits, retail chains

1. Employee Churn

Predict which employees are likely to leave and develop strategies to retain talent

Example: HR analytics in consulting firms

2. Strategic Investment Decisions

Use market forecasts to guide investment timing and location

Example: Expansion of electric vehicle infrastructure

3. Improving Advertisement Effectiveness

Identify which components of an ad drive conversions

Example: A/B testing by Airbnb

4. Product Recommendations

Forecast customer preferences based on past behavior

Example: Amazon, Netflix recommendation systems

Learning to Predict Real World Scenarios: Predicting Bus Ridership in NYC

Final Project Context

- The NYC Department of Transportation is forecasting number of customers to optimize resources:
 - How many buses are needed on a given day?
 - How many drivers should be scheduled?

Project Goal

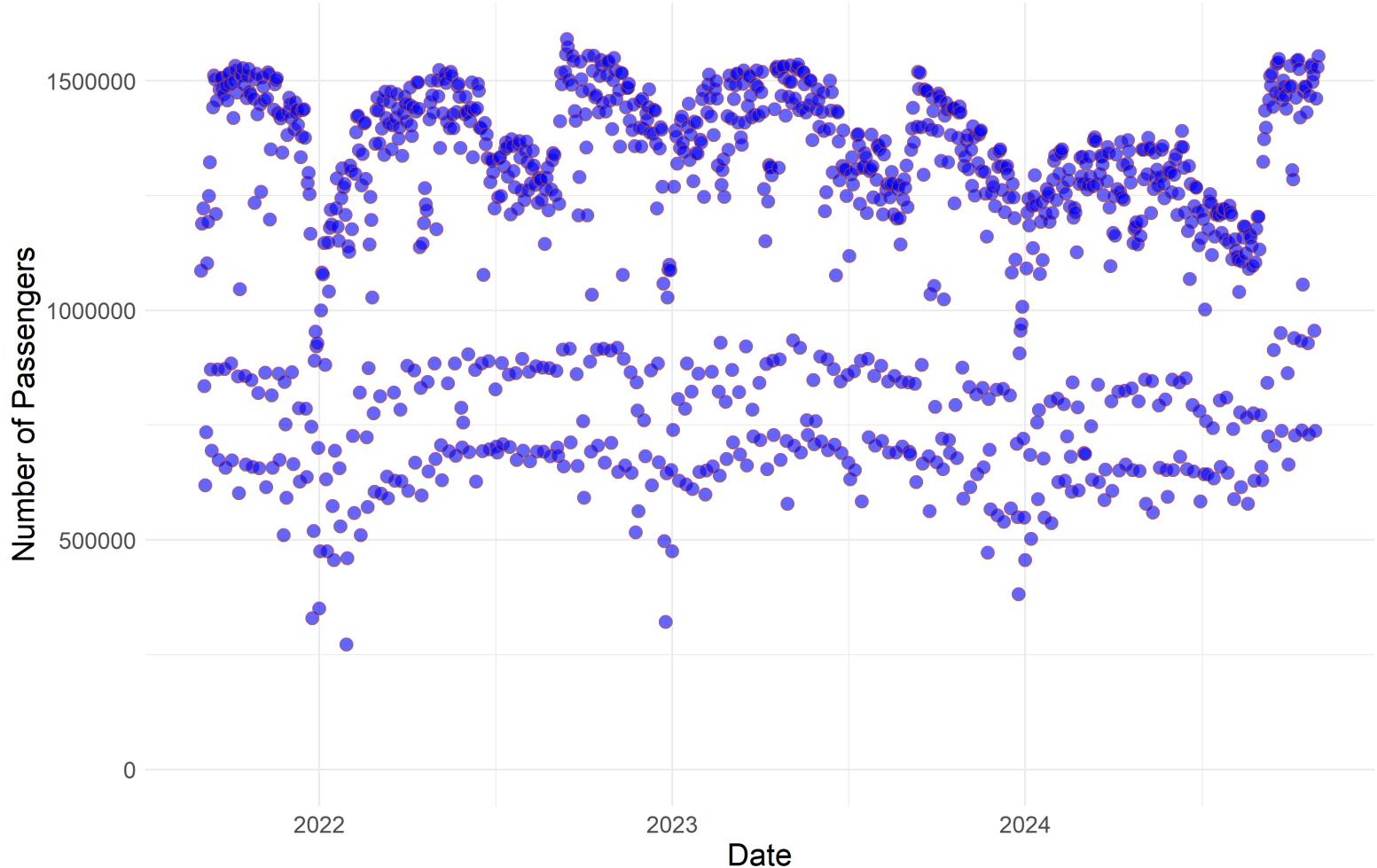
- Predict the **number of passengers using buses in NYC on December 3rd** (after the submission day).

Data Insights

- Rich quantitative data: Each payment generates a datapoint on customer bus usage.
- External factors affecting demand include:
 - School and holiday calendars
 - Weather conditions
 - City events
 - Road repairs

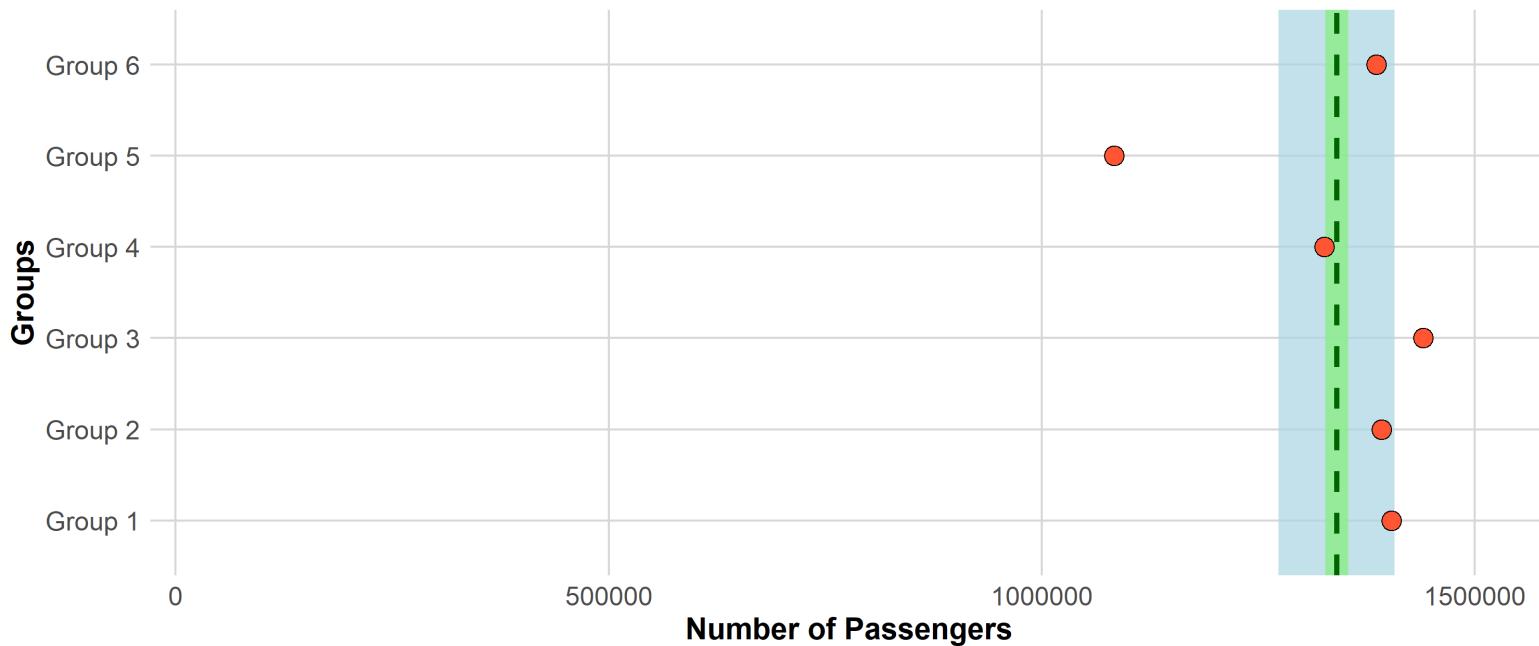
Daily Bus Ridership in NYC

Trends in passenger numbers over time



Predictions with 95% Confidence Intervals

Group-wise predictions with actual value as reference
Shaded areas: 2% (green) and 5% (blue) off actual value



Key Outcome

- Students built a **linear regression model** combining multiple data sources to generate forecasts.
- **Closest prediction was within 1% of the actual number of riders!**
- A reliable tool to improve resource allocation and decision-making.

Your turn

- Get in pairs
- Consider your past employment or your future employment
- Think about how forecasting could solve some problems in the context of industry you are considering
- (5 min)

What will you learn?

1. Getting Business Information from the Data
 - Analyze data, evidence, and arguments to make reasoned judgments
2. Problem Solving and Forecasting
 - Formulate, evaluate, and implement statistical models for business forecasting.
 - Interpret the results and validate assumptions
 - Key technical skill very valuable on the job market!
3. Decision Making and Communication
 - Choose optimal options to achieve objectives.
 - Communicate findings, conclusions, and recommendations effectively to business professionals

Organization

- **Lectures Schedule:**
 - Tuesday morning RH SA2
- **TA schedule:**
 - Weekly 1 hour lab session - timing/location to be determined
 - Office Hours with TAs - timing/location to be determined
 - Canvas Discussion Board
- **Textbooks**
 - See canvas course materials and syllabus
- **Organization**
 - see the website

Grading

- **70% In-Class Exams & Assignments**
 - 3 total exams (including final)
 - Combination of applied (in computer lab) and paper-based formats
- **20% Projects & Presentations**
 - 3 group-based take-home projects
 - Analyze data, write a short summary
 - Present findings briefly in class
- **5% In-Class Assignments**
 - 2 group forecasting competitions using real-world data
 - 2 individual theoretical quizzes (1 cheat sheet allowed)
 - **Lowest score among these 4 assignments will be dropped**
- **5% Participation & Engagement**
 - 2.5%: Flipped tasks and short homeworks (graded for completion)
 - 2.5%: Preparedness checks at the beginning of class
 - Random students selected in the first 5 minutes
 - Zero credit if absent or unprepared

Pre-requisites

- Mathematics III or Linear Algebra I
- Statistics II or Statistical Inference

Language

- Class is in English
- But your English is not evaluated
- If you don't know a word, feel free to ask
- Great occasion to learn vocabulary useful for interviews

ChatGPT & AI Tools Policy

- You are allowed to use ChatGPT and other language models.
 - Especially useful for refining code, summarizing concepts, and improving writing in projects.
- But be careful — copying output directly is plagiarism.
 - You must **understand and be able to explain** anything you include in your work.
 - During presentations, you will be expected to explain your project. If you can't, you'll receive **zero credit** for that assignment.
- ChatGPT is not allowed during exams.
- Use AI tools wisely — to learn, not to outsource thinking.
- Recommended for learning and review. Try asking:
 - "Why can't we say a 95% confidence interval contains the true value with 95% probability?"
 - "What's the relationship between Type I and Type II errors?"
 - "What is a p-value and what is it not?"

Explain p-value to me as Taylor Swift



Alright, let's do this the Taylor Swift way!

🎶 Hey, it's me, Taylor Swift! Let's talk about p-values like we're writing a song together. 🎶

So, imagine you're writing a song about someone who you *think* might be into you. But you're not sure, so you're looking for signs—like how often they text you, or if they show up at your shows. Your null hypothesis (just a fancy way of saying your starting assumption) is, "They're just being nice; they don't really like me."

Now, let's say you collect a bunch of "data" (you know, how often they text, how long they stay at your shows, whether they're in the front row singing every lyric). You run the numbers, and out comes the p-value.

If the p-value is super low (like 0.01), that's like saying, "There's only a 1% chance that all these signs are just a coincidence." In Taylor Swift terms, this is the part in the song where you're like, "*Oh my God, it's fate! They totally like me!*" So you reject the null hypothesis and start planning your next big love song.

But if the p-value is high (like 0.5), that's more like saying, "There's a 50/50 chance they're just being nice and it's nothing special." In this case, you might be singing something more like "*We are never ever getting back together*" because there's just not enough evidence to believe they're into you.

In short, the p-value helps you figure out whether to go all-in on those feelings or if you should just shake it off!

Academic Integrity

- It disrespects those supporting your education — you are not learning anything.
- It is unfair to your peers who put in genuine effort.
- It disrespects the professor and the academic environment.
- **Cheating is easy to detect:**
 - Quiz and exam questions are randomized.
 - A LockDown Browser will be used during quizzes.
 - Accessing quizzes outside the classroom without prior approval = **0 credit**.
 - Canvas access logs will be reviewed during assessments.
 - In-class assignments are recorded.
 - Anonymous reporting of cheating is available.
- **Penalties are severe:**
 - If caught cheating:
 - You will be reported to the administration.
 - You will automatically fail the course and must retake it.
 - On a second offense, **you will be expelled**.
 - Cheating has been reported in this course before.

Software: R

Performing Real Forecasts with Real Data

- **Widely Used**

R is one of the most popular languages for data analysis, statistical modeling, and forecasting — especially in academia and industry.

- **Free and Open-Source**

R is completely free to use and supported by a large global community.

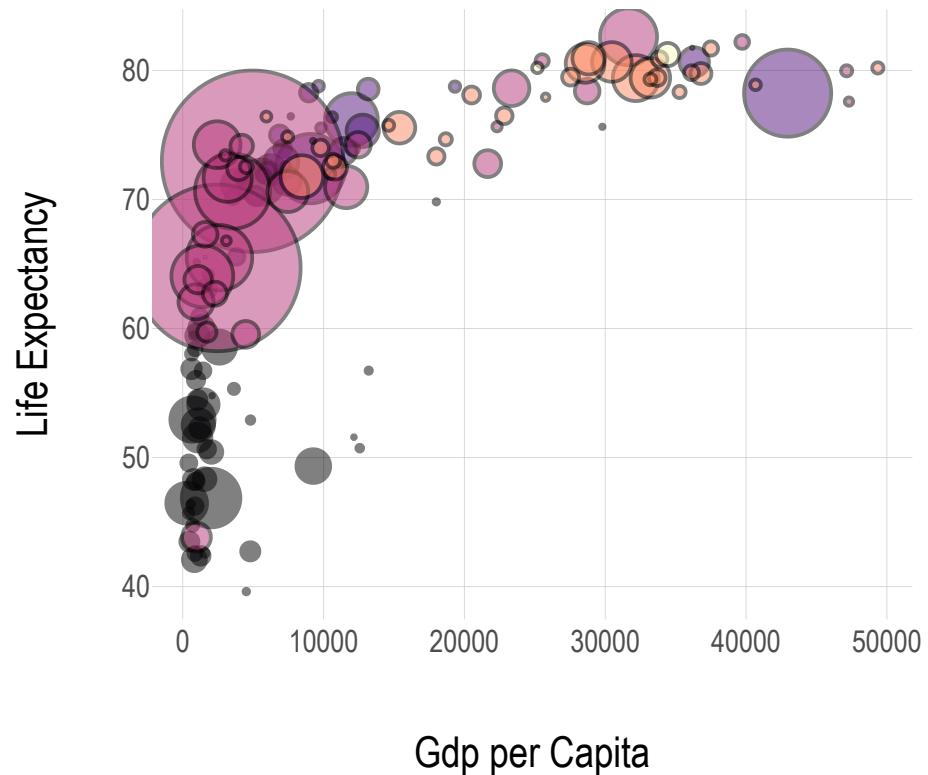
- **Great Community Support**

Thousands of tutorials, packages, and forums are available — it's easy to find help when you're stuck.

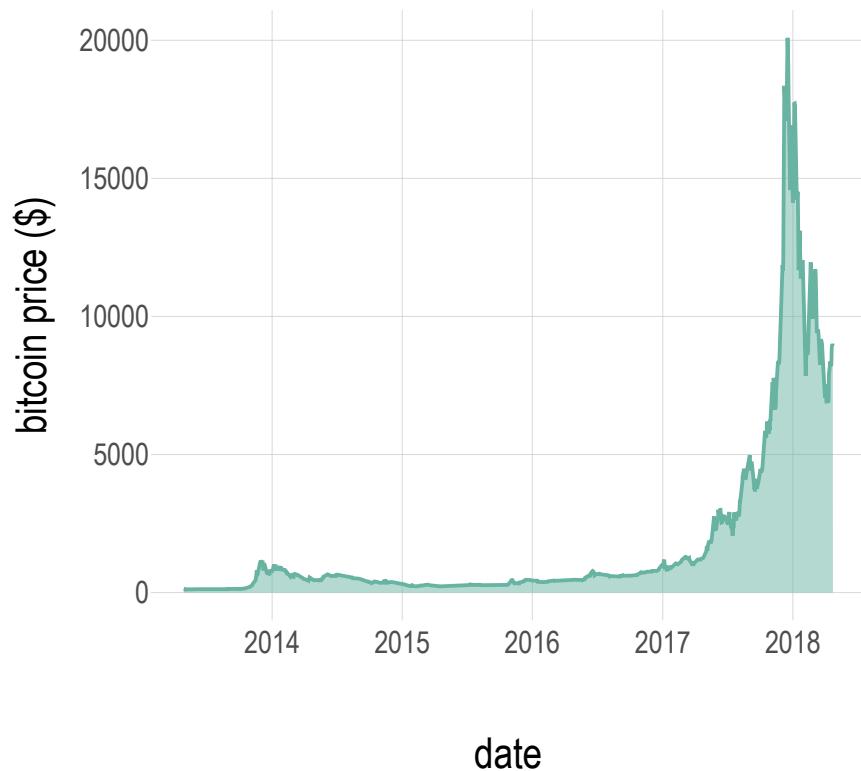
- **Powerful Data Visualization**

R includes powerful tools (like `ggplot2`) to create clear, publication-quality visualizations that help you communicate your insights effectively.

Life expectation vs GDP per Capita



Bitcoin price in time



Remarks

- Knowledge of theoretical statistics does not matter much if you can't apply it using modern tools
- We will use it for practical exercises with data
- You will use it for the final project
- You will use it for the applied exams
- We will learn some of it together in class
- TA lab sessions will further help with this
- Chatgpt is your friend

Introduction to Forecasting

Tools will often depend on the horizon and data availability

Forecasting Horizons

- **Very Short Horizon:**
 - High-Frequency Trading: Real-time price predictions for financial trading
- **Short Horizon:**
 - Public bikes: Predict the availability of bikes at bike station and adjust the number
- **Long Horizon:**
 - New obesity drug: Forecasting number of potential patients and their resources

Overview of Forecasting Techniques

1. Qualitative Forecasting

- Based on subjective judgment and expert opinions
- Suitable for unique situations or new markets
- Examples: Predicting economic impacts of oil price changes or political stability in a region

2. Quantitative Forecasting

- Uses historical data and numerical techniques
- Suitable when data is available and continuity assumptions hold true
 - *Continuity assumption:* past trends and relationships continue in the future
 - When it holds?
 - Interest rates and investments
 - When it does not hold?
 - Covid Cases & Deaths and Vaccines

Methods of Quantitative Forecasting

Time Series Forecasting

- **Time series data:** collection of data points for a single unit (one firm, one person, one country) ordered chronologically. Can be one or more variables.
- **Time series forecasting:** identifying patterns and trends in historical data to predict future values

Show 5 entries

Year	GDP
2010	9400
2011	10342
2012	10376
2013	10866
2014	11076

Showing 1 to 5 of 18 entries

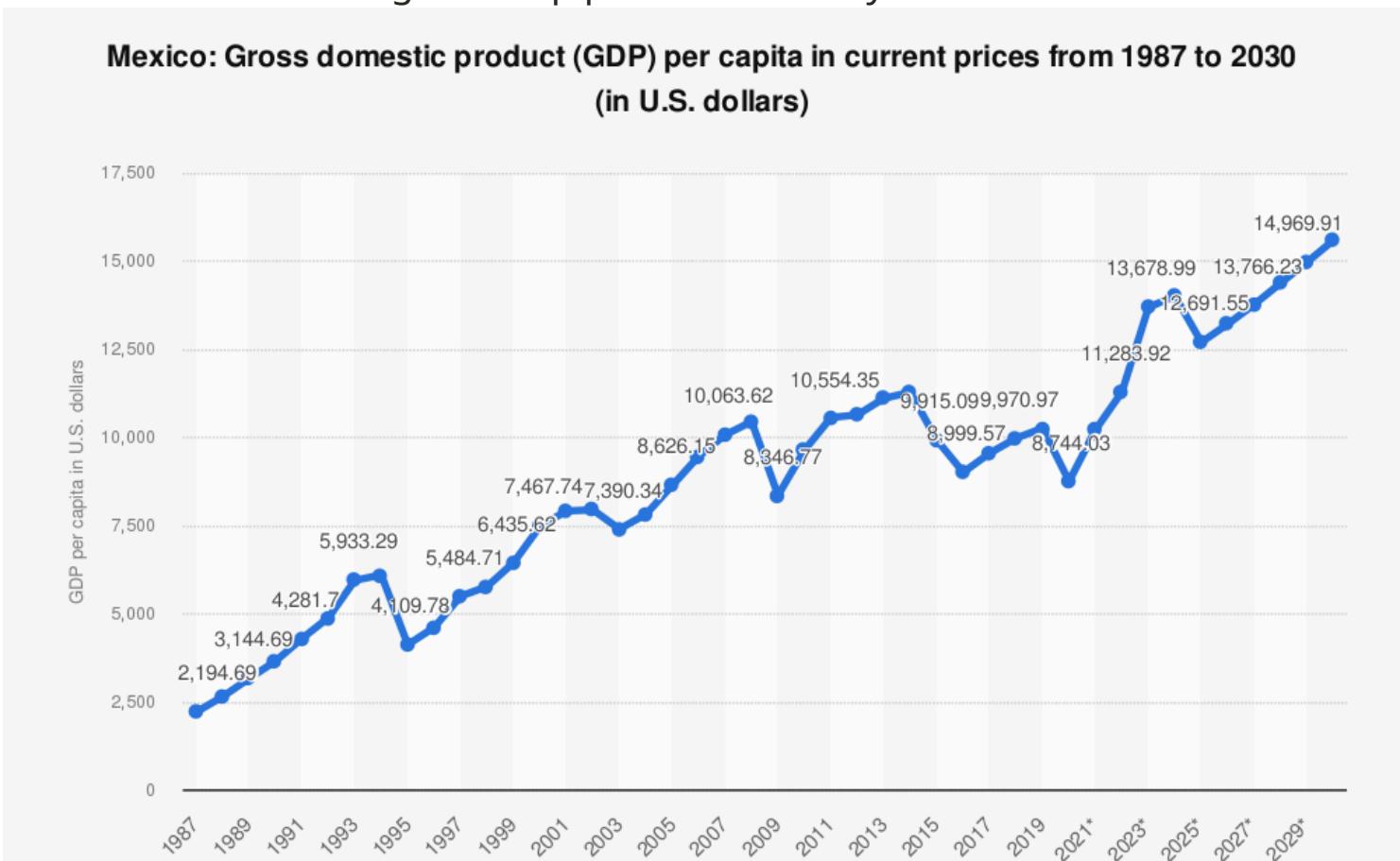
Previous 1 2 3 4 Next

In simple terms:

- We don't care about what causes what
- We just hope that past values of the variable and its historical behavior can predict its future values

Example: Forecasting of GDP

- Time series forecasting can help predict a country's Gross Domestic Product.



Explanatory Models

- We have data on both the variable of interest and other variables related to it
- We consider how other variables impact the outcome of interest
- We use these relationships to make forecasts

Example: Sales at a new location

- Should we open a new Starbucks at ITAM?
- Using existing locations, analyze impact of:
 - Foot traffic
 - Neighborhood income
 - Competitors' stores
- Given these relationships, what would be sales at ITAM?

Sales Data Analysis						
S_ID	Address	F_Traffic	N_Income	C_Stores	Sales	
1	123 Main St	100	60000	2	5000	
2	456 Oak Ave	150	75000	1	6000	
3	789 Elm Rd	80	55000	3	4500	
4	321 Maple Ln	120	80000	2	7000	
5	543 Birch St	200	70000	1	8000	
6	876 Pine Ave	90	60000	3	5500	
7	987 Cedar Rd	180	90000	2	9000	
8	654 Walnut Ln	110	85000	1	7500	
9	234 Spruce St	140	75000	2	6500	
10	ITAM	170	65000	1		

Showing 1 to 10 of 10 entries

Previous 1 Next

Steps of Forecasting

1. Problem Definition

- Clearly define the forecasting objective.

2. Gather Data

- Identify and collect relevant data.

3. Preliminary Explanatory Analysis

- Understand data characteristics and relationships.

1. Choosing and Fitting the Model

- Select and fit the appropriate forecasting model.

4. Evaluating the Model

- Assess the model's performance using historical data.

5. Communicate the results

- Visualize and interpret the outcomes and write a report or presentation

Remainder of the course

Quantitative Forecasting

1. Ungraded quiz
2. Review of Statistics
3. Simple linear regression
4. Multiple linear regression
5. Time Series

Questions?

