Class 1: Introduction to Business Forecasting

Krzysztof Zaremba

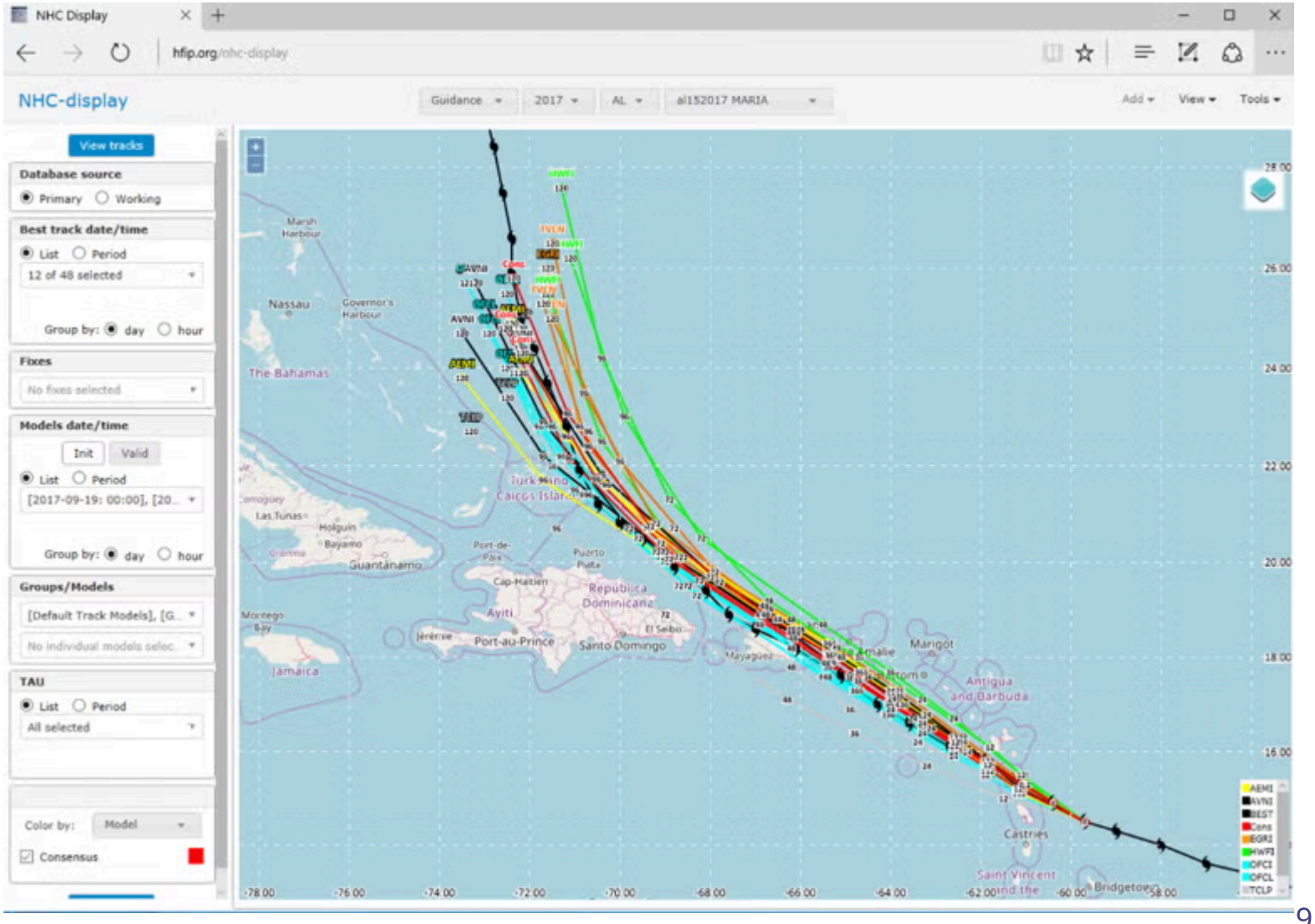
Who am I?

- Krzysztof Zaremba
  - **Education**: PhD in Economics from Columbia University
  - **Research focus**: Applied Econometrics, Health Economics
  - **Email:** zaremba@itam.mx
  - **Office:** In Santa Teresa Campus
  - **Office hours:** on zoom, Monday 2pm-3pm, email me if different time needed

What Is This Class About?

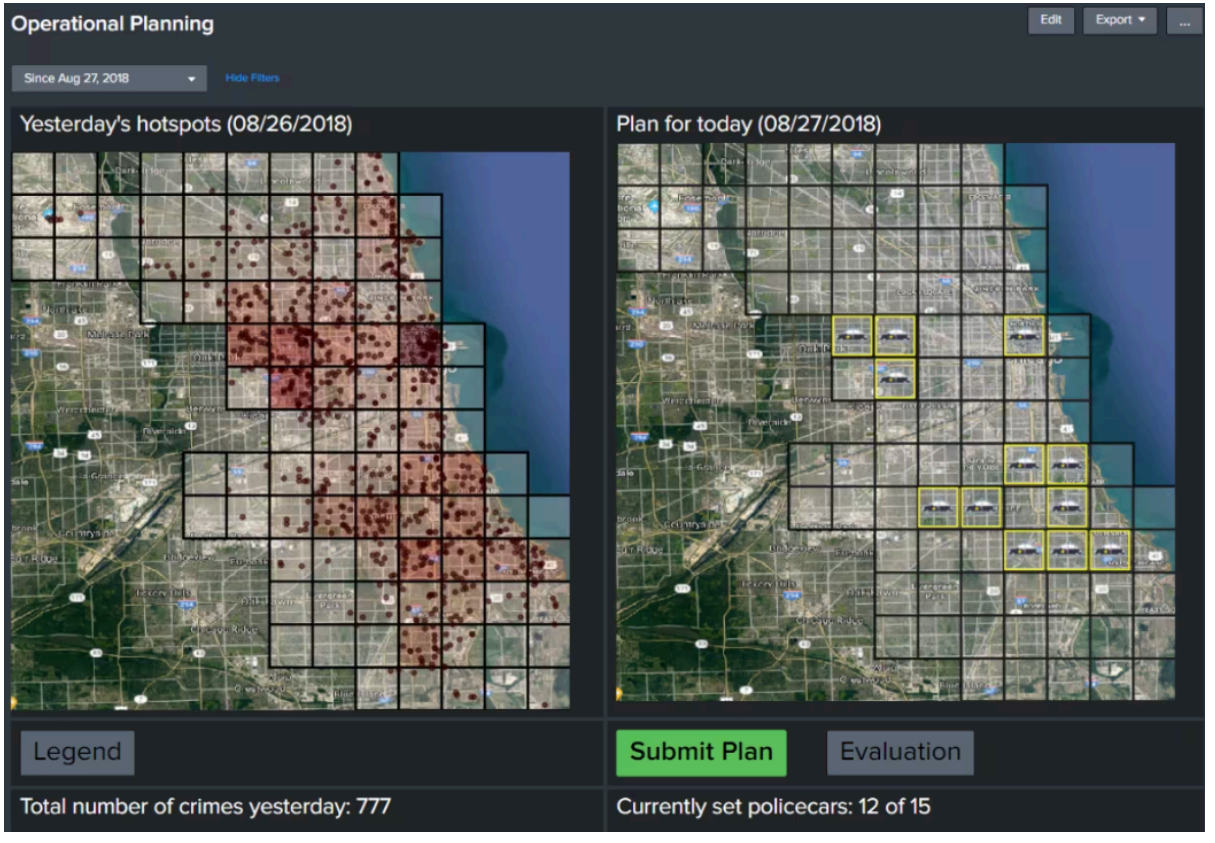
Forecasting involves making predictions about future events based on historical data and relevant information

1. **Weather Forecasting**
   - Predicting hurricanes → timely reactions and preparations.



1. **Epidemic Forecasting**
    - Predicting disease outbreaks → effective public health responses.

1. **Criminal Activity Forecasting**
   - Identifying crime hot-spots → increase the number of patrols.
   - Identifying individuals at risk of committing crimes → sending warning letters.



What Is This Class About?

Forecasting in Business

- Improve decision making
- With increasing availability of data, firms rely on almost scientific methods to make decisions

Applications

1. **Pricing**: Optimize pricing (Example: Health Insurance, Uber)
2. **Employee Churn**: Identify factors predicting employees leaving the company and implement measures to retain valuable talent (Example: EC)
3. **Anticipating Demand**: Adjust inventory management to minimize waste and meet customer needs effectively (Example: Meal Kits)
4. **Strategic Investment Decisions**: Forecast market trends to predict suitability for investments (Example: Electric Cars)

5. **Improving Advertizement**: Idenfitying which elements of the ad are most effective (Example: AirBnB)

6. **Suggesting Products**: Recommending products to customers based on their past behavior (Example: Amazon, Netflix)

Learning to Predict Real World Scenarios: Predicting Demand for Public Transportation in New York

Final Project Context

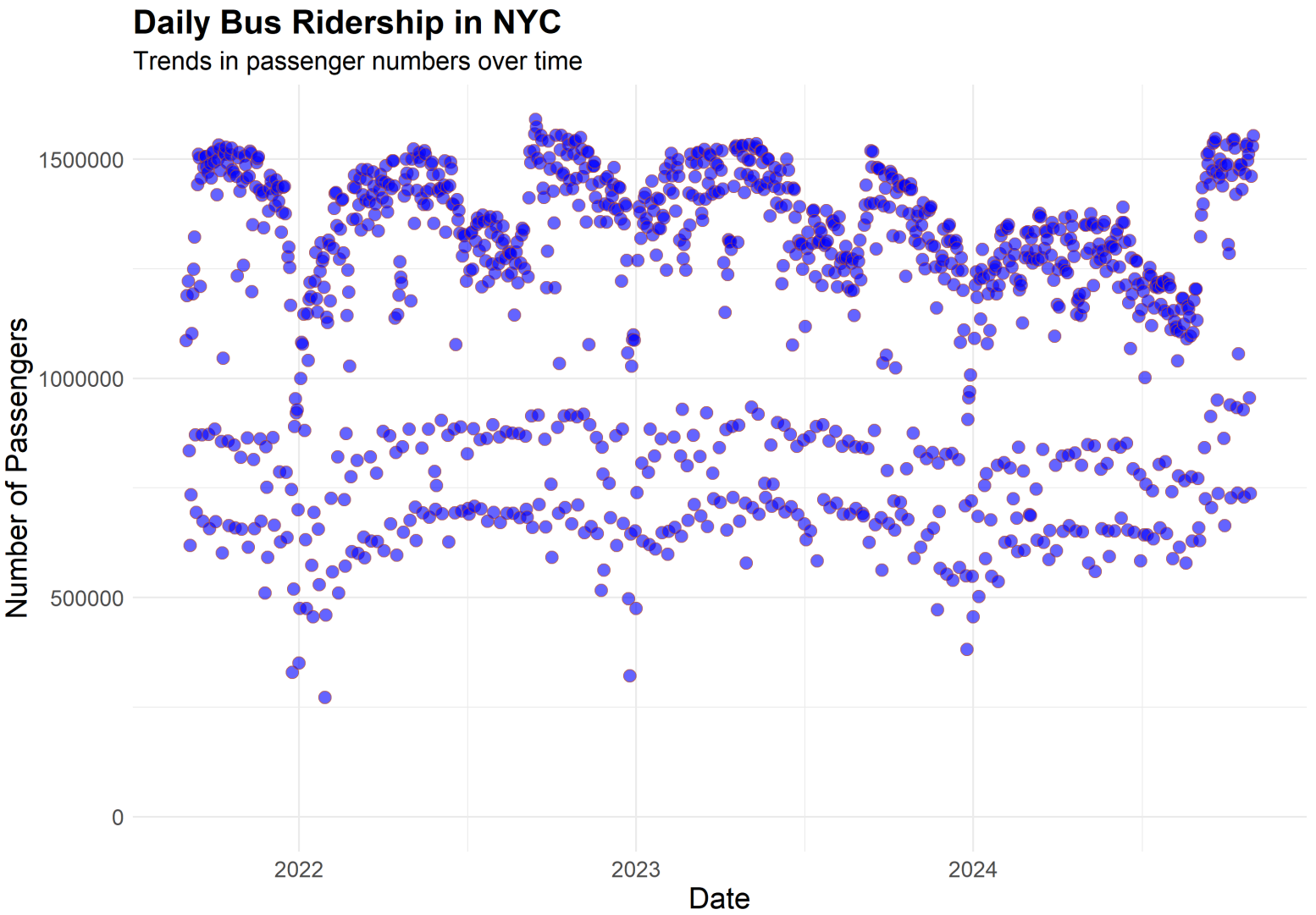
- The NYC Department of Transportation is forecasting number of customers to optimize resources:
  - How many buses are needed on a given day?
  - How many drivers should be scheduled?

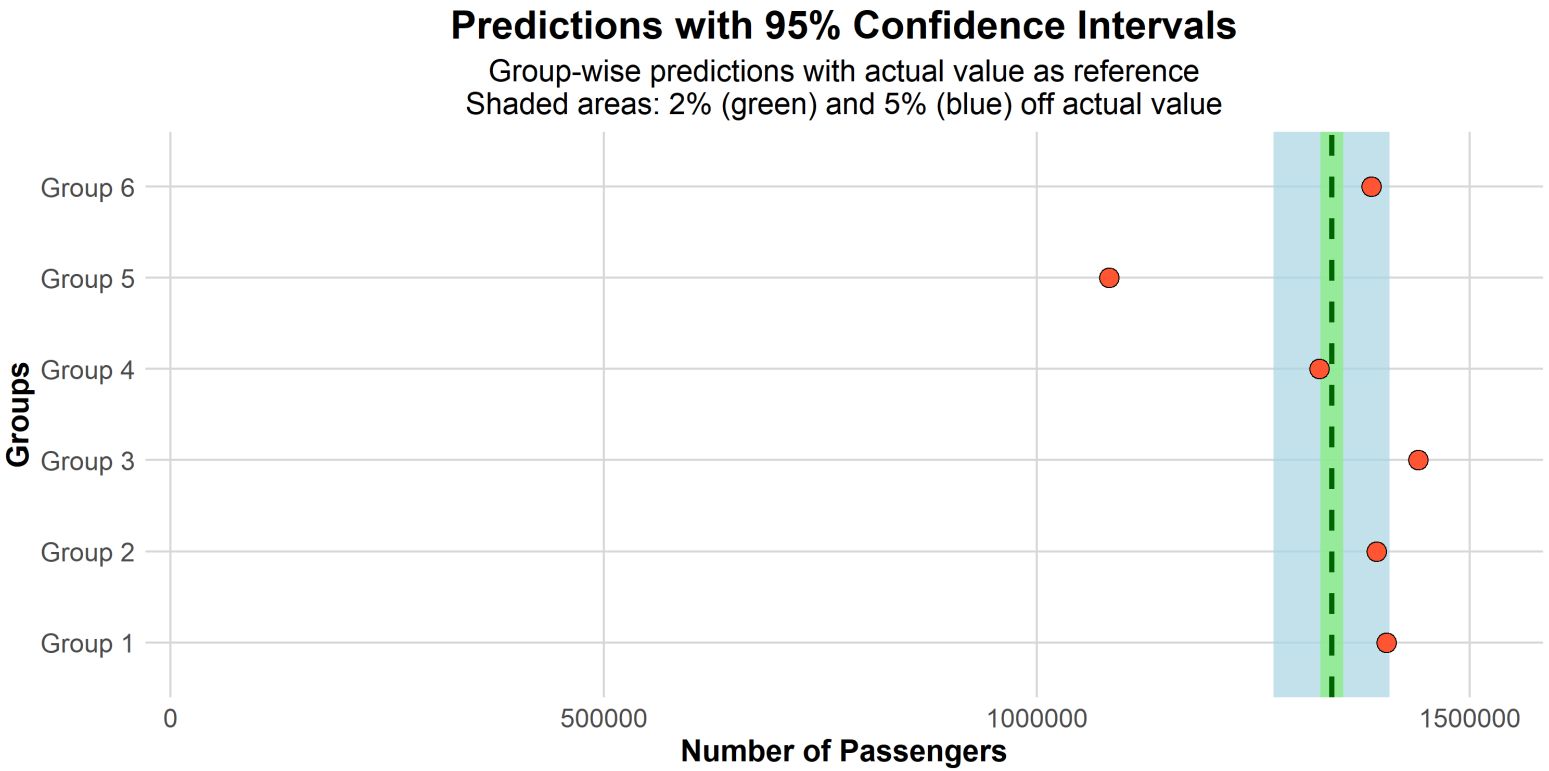
Project Goal

- 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
Number of Passengers
1500000
1000000
500000
0
2022
2023
2024
Date



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 & Thursday morning RH308
- **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

- 20% 2x Midterms
- 30% Final Exam - you need at east 60 to pass the class
- 10% Quizzes
  - *Type 1 Quiz*: In-class, individual, theory + some calculations, 30 min max, one week notice, around 3 per semester. You can bring one page of handwritten notes. Printed or electronic notes will not be accepted.
  - *Type 2 Quiz*: At home, individually or in pairs, practice with data, 24h to complete, 2 per semester, open book.
  - Lowest grade among all is dropped
- 15% Final Project (In groups of 3)
- 5% Homework
  - Graded on completion, not correctness, but you need to show an attempt to solve
  - You can skip one

Last Semester Median Grades

- Midterm 1: 70
- Midterm 2: 57
- Final Exam: 62.5
- Quiz: 81
- Final Project: 84
- Final Grade: 68.5

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
- Exams will be available in both languages but you have to answer in English
- I will understand your English
- Great occasion to learn vocabulary useful for interviews

ChatGPT Policy

- We are not trying to fight the technology
- You are allowed to use ChatGPT/other language models for your advantage
    - You can use it for at home quizzes and final project to improve your code and writing, but...
    - Just copying directly the ouput from CHATGPT is plagiarism
    - I sometimes ask to explain the code you submitted, so you need to understand everything you put in an assignment
- Use it wisely to learn, and then inform your own answers with its ouput
- I recommend it for learning:
    - Try typing in:
    - Why we cannot say that the confidence interval contains the true value with 95%
    - What the relationship between type 1 and type 2 errors
    - What is a p-value?

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!

Integrity

- Cheating is disrespectful for your study sponsors, you are not learning anything, and you risk getting kicked out
- Cheating is disrespectful to your peer students, who make effort in the class
- Cheating is disrespectful to the professor

- **Cheating is easy to discover**:
  - Questions are usually randomized
  - We will use lockdown browser for the quizes
  - If you access the quiz in a place other than the classroom, you get 0 (unless agreed beforehand)
- **Penalties are very strong**
  - I take cheating extremely seriously
  - When discovered cheating, I will report you to the administration. You will fail the class and have to retake it. At the second time, you are expelled.
  - I did report cheating in this course before.

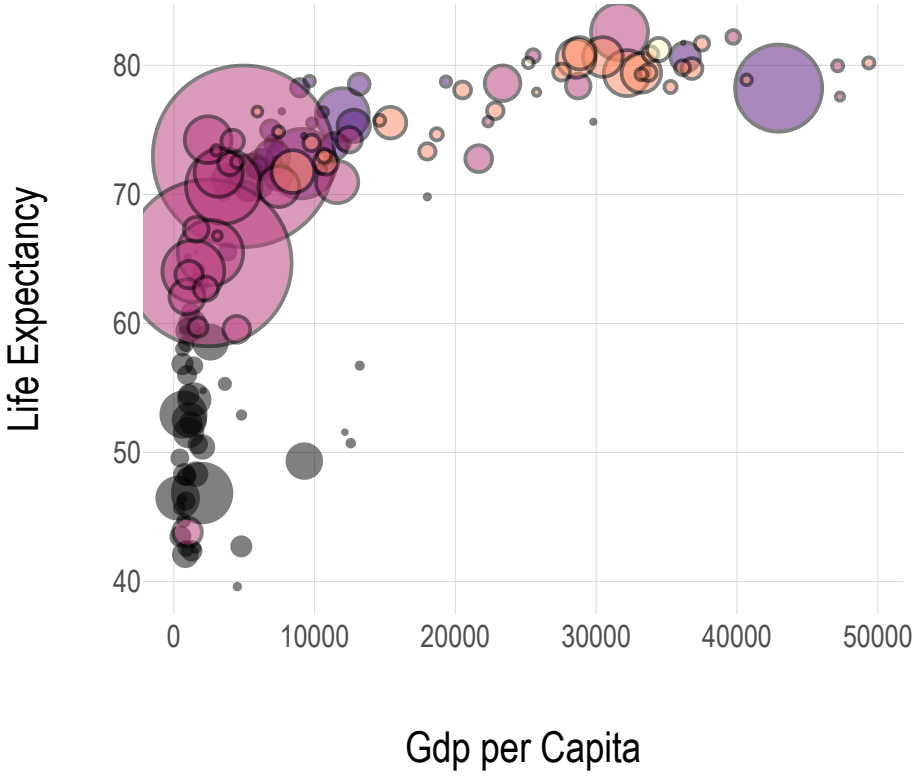
- If you are honest, you don't have to worry about all the above

Software - R

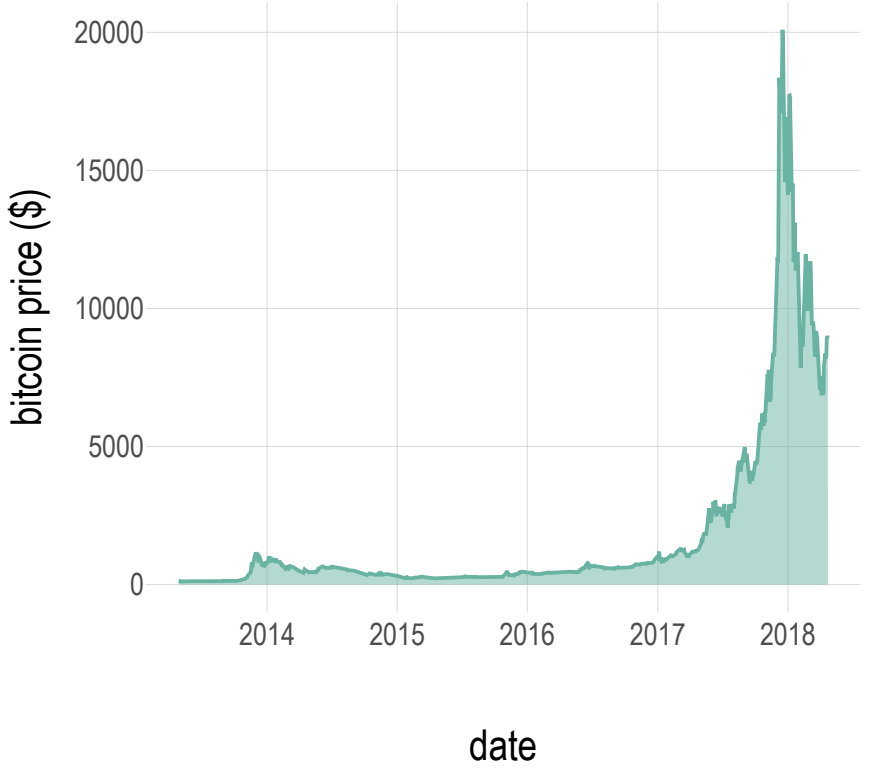
Performing actual forecasts with data

- Widely Used: R is a popular language for data analysis and statistical computing
- Open-Source: R is free!
- Community Support: You can easily find plenty of tutorials and help
- Data Visualization: Amazing vizualizations and ways to communicate your findings

Life expectation vs GDP per Capita



Bitcoin price in time



Remarks

- We will use it for practical exercises with data
- You will use it for the final project
- 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
  - Traffic Flow Management: Optimizing traffic signals based on real-time data
- Short Horizon:
  - Retail Store: Forecasting cashier scheduling based on historical data
  - 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
  - Investing in new country: Forecasting political stability

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

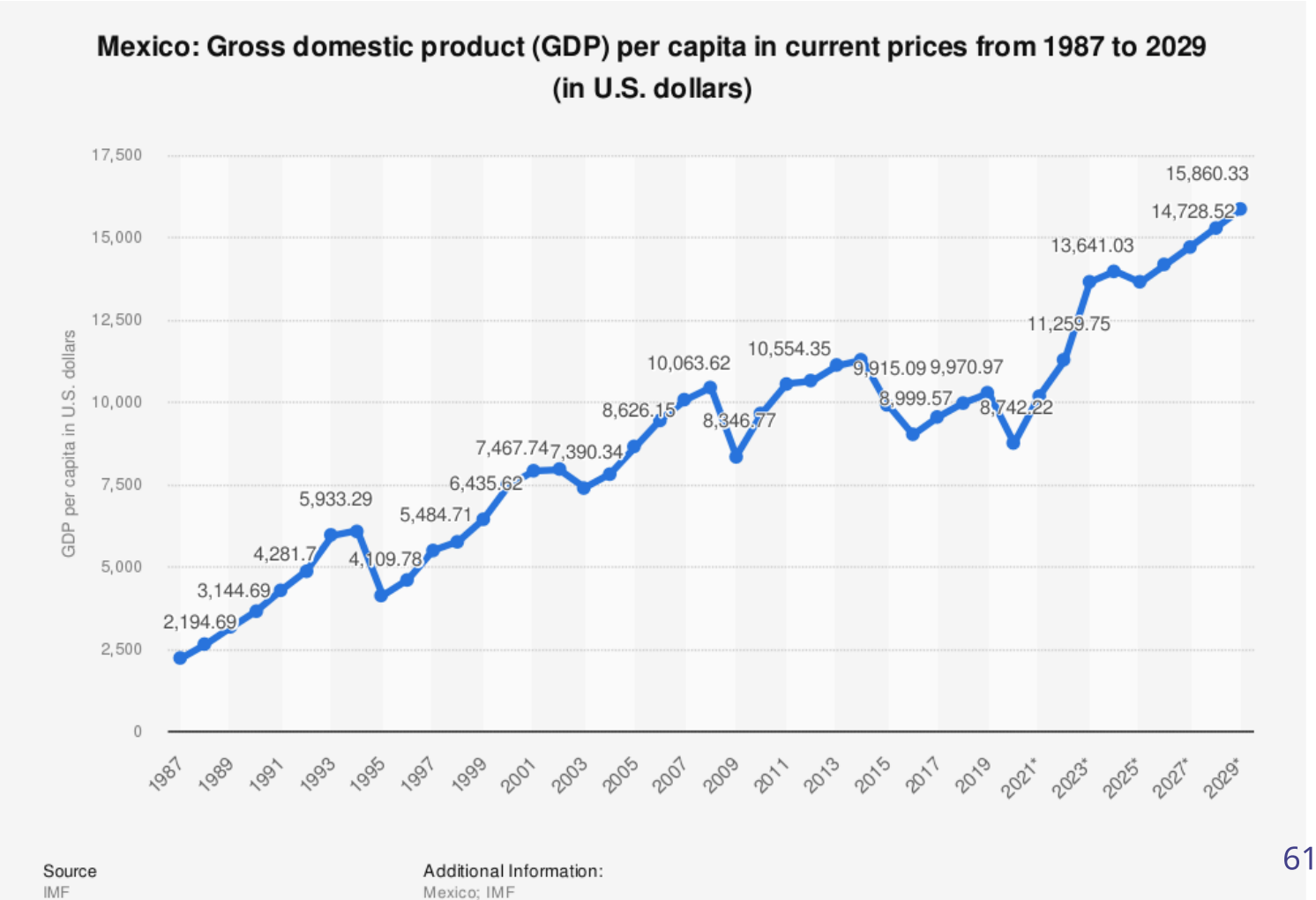
Previous 1 2 3 4 Next

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?

Show 10 entries

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.
   - Example: Forecasting ride demand during holidays for Uber.
2. **Gather Data**
   - Identify and collect relevant data.
   - Example: Historical ride data and local event information.
3. **Preliminary Explanatory Analysis**
   - Understand data characteristics and relationships.
   - Example: Analyzing trends and patterns in ride demand during holidays.
4. **Choosing and Fitting the Model**
   - Select and fit the appropriate forecasting model.
   - Example: Estimate multiple linear regression or time series forecasting.
5. **Evaluating the Model**
   - Assess the model's performance using historical data.
   - Example: Comparing model predictions with actual ride demand during past holidays.

Methods of Qualitative Forecasting

Delphi Method

- A structured communication process to reach a consensus for complex, uncertain and long terms forecasting tasks
  1. Select a group of experts
  2. Invite them to the study. They are anonymous and don't talk to each other!
  3. Ask them to answer a questionnaire
  4. Get initial responses
  5. Compile them into summary
  6. Send them summary and get their feedback with refined answers
  7. Reiterate until consensus is reached or no further improvement

Example: Determining AI threats

- What are the risks of AI developments?
- Panel of experts from academia and industry
  - Computer scientists, engineers, CEOs of AI companies, ethic experts
- Send them questionnaires asking about potential threats
- Compile responses into summary and send them back
- Get more rounds of responses until consensus
- Identify the most probable risks

Brainstorming

- Creative technique for generating ideas.
- Encourages free thinking and building on suggestions.
- Appropriate for exploring possibilities.
  - Form a group (no need for experts)
  - State the problem
  - Encourage ideas, no matter how crazy
  - Build and combine each others' ideas
  - Document the ideas and synthesize them

Example: Enhancing Employee Engagement

- Tech company's HR department.
- Representatives from HR, IT, and different departments.
- Generate ideas for a mobile app to enhance employee engagement.
- Write them down and implement the relevant ones

Panel of Experts

- Assemble knowledgeable individuals
  - At the same time and spot
- They meet, offer insights and expertise, and discuss
- Aid in well-informed decisions.
- Sometimes ends up with a report with conculsions

Example: Environmental Policy Formulation

- Government agency want to find identify and address most pressing environmental issues
- Environmental scientists, economists, conservationists, and policymakers.
- Discuss policy options.
- Create comprehensive environmental policies.

Focus Groups

- Gather diverse participant - not necessarily experts
- Share perceptions, attitudes, and opinions.
- Provide qualitative data and consumer insights.

Example: Market Research for a New TV SHOW

- Proposing a new TV Show and trying to see how well it will do
- Participants from various demographics.
- Understand consumers' preferences and perceptions about the TV show
- Fine-tune the product and marketing strategy.

Remainder of the course

Quantatitve Forecasting

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

Fundamental tools to:

- Make predictions,
- Quantify uncertainty,
- Intepret it and communicate it

Questions?