# Course Overview

Ling 250/450: Data Science for Linguistics

C.M. Downey

Spring 2026

# What is Data Science anyway?

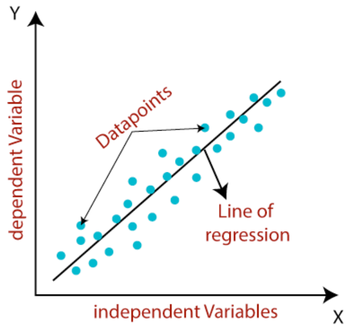
# What is Data Science?

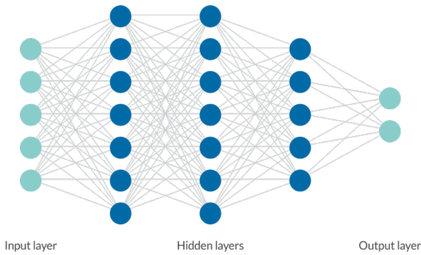
# What is Data Science?

- Data Science is (in my opinion) somewhat **poorly defined**

# What is Data Science?

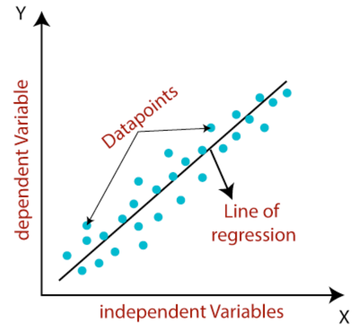
- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:

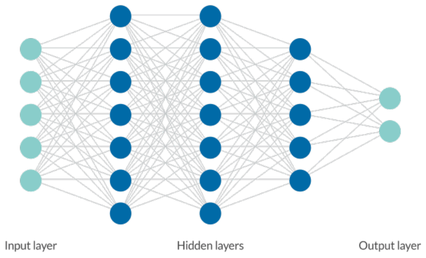



# What is Data Science?

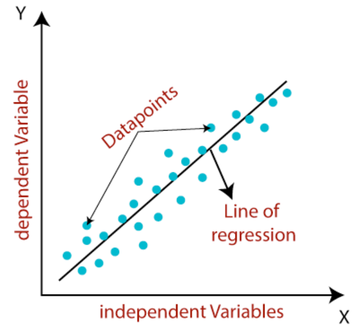
- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:
  - **Collecting**, **manipulating**, and **cleaning** data

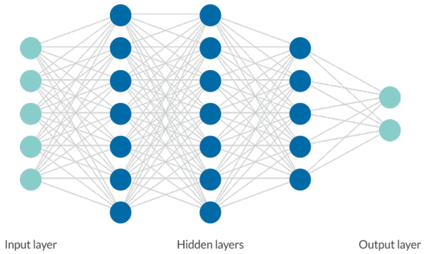



# What is Data Science?

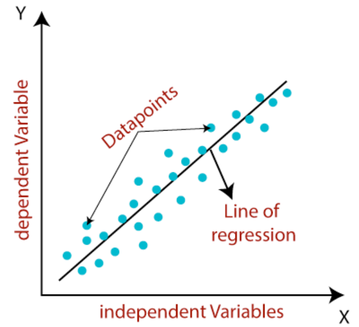
- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:
  - **Collecting**, **manipulating**, and **cleaning** data
  - Running **statistical tests** and models

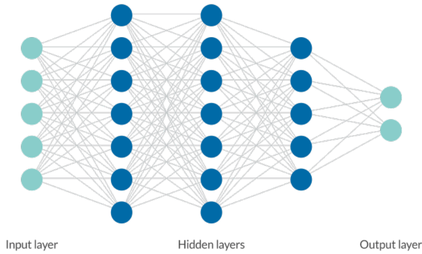



# What is Data Science?

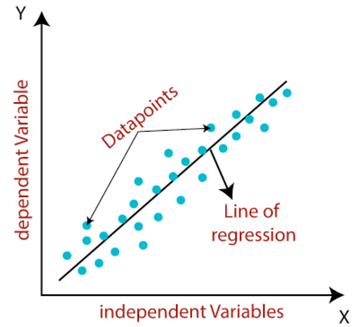
- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:
  - **Collecting**, **manipulating**, and **cleaning** data
  - Running **statistical tests** and models
  - Training and using **machine learning** models

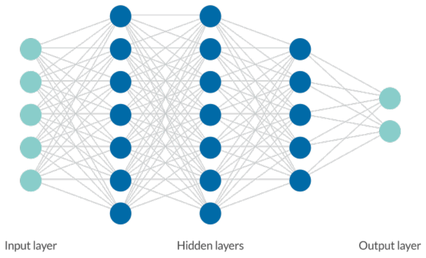



# What is Data Science?

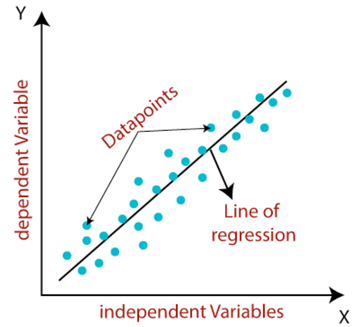
- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:
  - **Collecting**, **manipulating**, and **cleaning** data
  - Running **statistical tests** and models
  - Training and using **machine learning** models
  - **Interpreting** meaning or trends from data

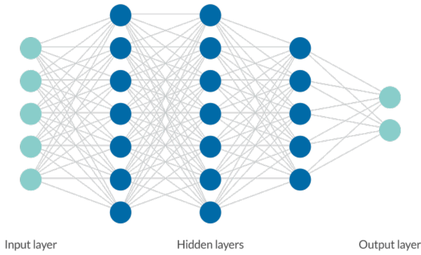



# What is Data Science?

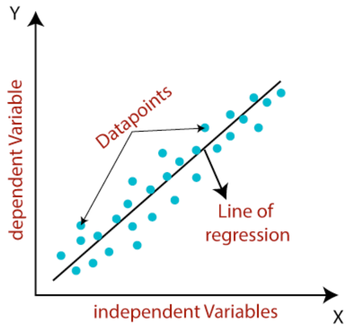
- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:
  - **Collecting**, **manipulating**, and **cleaning** data
  - Running **statistical tests** and models
  - Training and using **machine learning** models
  - **Interpreting** meaning or trends from data
  - **Visualizing** data in useful and novel ways

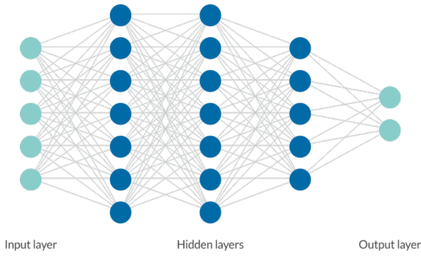



# What is Data Science?

- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:
  - **Collecting**, **manipulating**, and **cleaning** data
  - Running **statistical tests** and models
  - Training and using **machine learning** models
  - **Interpreting** meaning or trends from data
  - **Visualizing** data in useful and novel ways

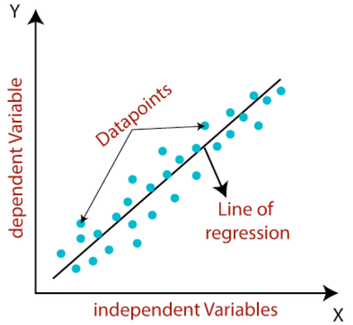


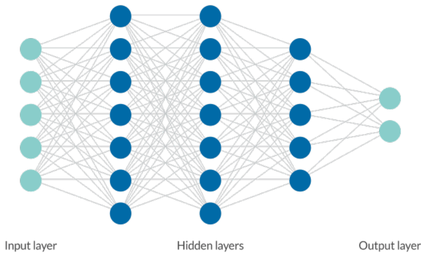




- All of these skills are used in **many other fields!** (And most pre-date “Data Science”)

# What is Data Science?

- Data Science is (in my opinion) somewhat **poorly defined**
- Things people will often refer to as Data Science:
  - **Collecting**, **manipulating**, and **cleaning** data
  - Running **statistical tests** and models
  - Training and using **machine learning** models
  - **Interpreting** meaning or trends from data
  - **Visualizing** data in useful and novel ways




- All of these skills are used in **many other fields!** (And most pre-date “Data Science”)
- Data Scientists might only specialize in **one or a few** of these skills

# Is it actually science?

# Is it actually science?


- Again in my opinion, “Data Science” might be a **bit of a misnomer**

# Is it actually science?

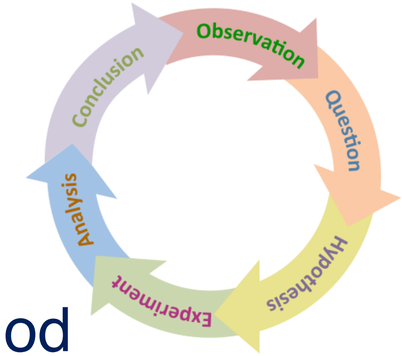

- Again in my opinion, “Data Science” might be a **bit of a misnomer**
- These techniques are **often** but **not always** applied to science
  - An important distinction is that science follows the **Scientific Method**, and mostly acquires knowledge through **hypothesis testing**
  - Most of the hallmarks of “Data Science” can be used **without** following the Scientific Method

# Is it actually science?


- Again in my opinion, “Data Science” might be a **bit of a misnomer**
- These techniques are **often** but **not always** applied to science
  - An important distinction is that science follows the **Scientific Method**, and mostly acquires knowledge through **hypothesis testing**
  - Most of the hallmarks of “Data Science” can be used **without** following the Scientific Method
- While the name “Data Science” is here to stay, it might (sometimes) be more accurate to talk about **data {analysis, exploration, engineering}**

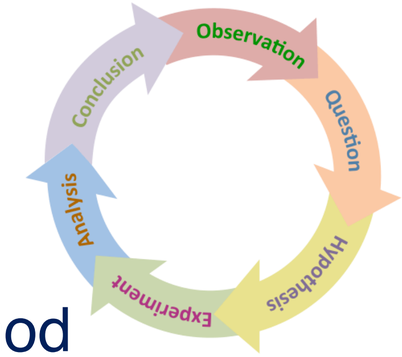
# “Data Science” and Science

# “Data Science” and Science



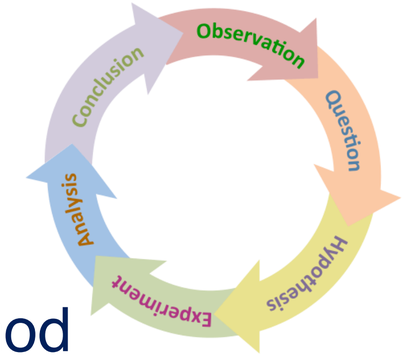

- Data Science techniques can be an integral part of the Scientific Method

# “Data Science” and Science



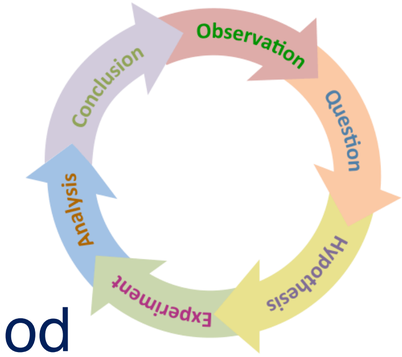

- Data Science techniques can be an integral part of the Scientific Method
  - Prior to experimentation: **explore**, **analyze**, and **visualize** patterns in data and previous experimental results in order to form a new **testable hypothesis**

# “Data Science” and Science



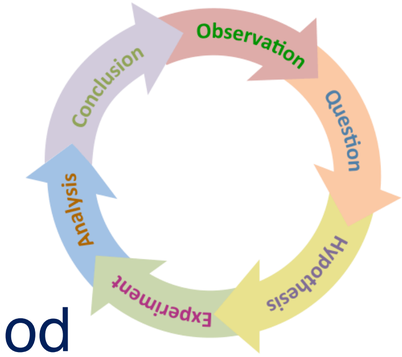

- Data Science techniques can be an integral part of the Scientific Method
  - Prior to experimentation: **explore**, **analyze**, and **visualize** patterns in data and previous experimental results in order to form a new **testable hypothesis**
  - During experimentation: **collect**, **clean**, and **organize** data that allows you to properly **test your hypothesis**, ideally with **statistical methods**

# “Data Science” and Science




- Data Science techniques can be an integral part of the Scientific Method
  - Prior to experimentation: **explore**, **analyze**, and **visualize** patterns in data and previous experimental results in order to form a new **testable hypothesis**
  - During experimentation: **collect**, **clean**, and **organize** data that allows you to properly **test your hypothesis**, ideally with **statistical methods**
  - After experimentation: **analyze** results to tell if your hypothesis is supported. Create **visualizations** to share your results. **Organize** your data for dissemination and experimental **replication**

# “Data Science” and Science




- Data Science techniques can be an integral part of the Scientific Method
  - Prior to experimentation: **explore**, **analyze**, and **visualize** patterns in data and previous experimental results in order to form a new **testable hypothesis**
  - During experimentation: **collect**, **clean**, and **organize** data that allows you to properly **test your hypothesis**, ideally with **statistical methods**
  - After experimentation: **analyze** results to tell if your hypothesis is supported. Create **visualizations** to share your results. **Organize** your data for dissemination and experimental **replication**
- We will read later about the importance of **separating exploratory studies** and **hypothesis testing** (but we will engage in both in this course!)

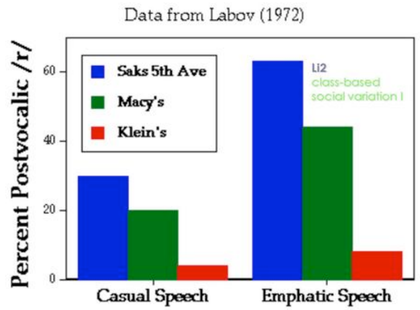
# Data Science and Linguistics

# Data Science and Linguistics

- Linguistics (mostly) acts as the **science of language**
  - Thus, DS can be used to **explore and test hypotheses** about language and how it works

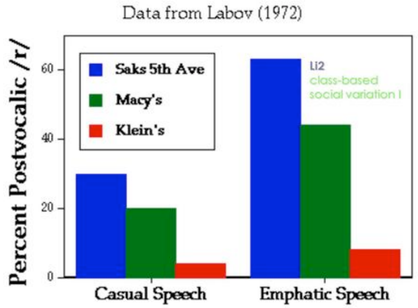
# Data Science and Linguistics

- Linguistics (mostly) acts as the **science of language**
  - Thus, DS can be used to **explore and test hypotheses** about language and how it works
- Some linguistic sub-fields have **long embraced DS methods**
  - e.g. Sociolinguistics, Psycholinguistics, and “Lab”/Experimental Phonetics



# Data Science and Linguistics

- Linguistics (mostly) acts as the **science of language**
  - Thus, DS can be used to **explore and test hypotheses** about language and how it works
- Some linguistic sub-fields have **long embraced DS methods**
  - e.g. Sociolinguistics, Psycholinguistics, and “Lab”/Experimental Phonetics
- Others have been historically **more resistant** (especially Chomsky)
  - “One’s ability to produce and recognize grammatical utterances is **not based on notions of statistical approximations** and the like” (Chomsky 1957)
  - Syntax and Generative Grammar have tended towards methods more similar to **math** and **philosophy** than (experimental) science




⟨SENTENCE⟩ → ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⟨NOUN-PHRASE⟩ → ⟨CMPLX-NOUN⟩ | ⟨CMPLX-NOUN⟩⟨PREP-PHRASE⟩
⟨VERB-PHRASE⟩ → ⟨CMPLX-VERB⟩ | ⟨CMPLX-VERB⟩⟨PREP-PHRASE⟩
⟨PREP-PHRASE⟩ → ⟨PREP⟩⟨CMPLX-NOUN⟩
⟨CMPLX-NOUN⟩ → ⟨ARTICLE⟩⟨NOUN⟩
⟨CMPLX-VERB⟩ → ⟨VERB⟩ | ⟨VERB⟩⟨NOUN-PHRASE⟩
⟨ARTICLE⟩ → a | the
⟨NOUN⟩ → boy | girl | flower
⟨VERB⟩ → touches | likes | sees
⟨PREP⟩ → with

# Rationalism and Empiricism in Linguistics

# Rationalism and Empiricism in Linguistics

- Chomsky's school of Linguistics uses **rationalist** techniques for insight on language, while experimental and data-driven methods are considered **empiricist**. (Very) roughly:

# Rationalism and Empiricism in Linguistics

- Chomsky's school of Linguistics uses **rationalist** techniques for insight on language, while experimental and data-driven methods are considered **empiricist**. (Very) roughly:
  - Rationalism: knowledge is obtained through **introspection** and **logical processes** like deduction and induction

# Rationalism and Empiricism in Linguistics

- Chomsky's school of Linguistics uses **rationalist** techniques for insight on language, while experimental and data-driven methods are considered **empiricist**. (Very) roughly:
  - Rationalism: knowledge is obtained through **introspection** and **logical processes** like deduction and induction
  - Empiricism: knowledge is obtained through observation of **external experience**, especially the results of **controlled experiments**

# Rationalism and Empiricism in Linguistics

- Chomsky's school of Linguistics uses **rationalist** techniques for insight on language, while experimental and data-driven methods are considered **empiricist**. (Very) roughly:
  - Rationalism: knowledge is obtained through **introspection** and **logical processes** like deduction and induction
  - Empiricism: knowledge is obtained through observation of **external experience**, especially the results of **controlled experiments**
- Note that these philosophies are **not mutually exclusive**, but researchers have historically tended to argue that one may be more reliable than the other

# Why be an empiricist linguist?

# Why be an empiricist linguist?

- Rationalist approaches to Linguistics often rely on the **introspective judgements** from just one or a few people
  - e.g. “is this sentence grammatical?”, “do these words mean the same thing?”

# Why be an empiricist linguist?

- Rationalist approaches to Linguistics often rely on the **introspective judgements** from just one or a few people
  - e.g. “is this sentence grammatical?”, “do these words mean the same thing?”
- Most linguists eventually have to confront the major insight of Sociolinguistics: **language varies extensively**
  - (Between people, communities, genres, conversation partners, etc.)
  - The same person might have different judgements at different times!

# Why be an empiricist linguist?

- Rationalist approaches to Linguistics often rely on the **introspective judgements** from just one or a few people
  - e.g. “is this sentence grammatical?”, “do these words mean the same thing?”
- Most linguists eventually have to confront the major insight of Sociolinguistics: **language varies extensively**
  - (Between people, communities, genres, conversation partners, etc.)
  - The same person might have different judgements at different times!
- This makes it questionable what the introspection of **any one person** can tell us about **language in general**

# Why be an empiricist linguist?

- Rationalist approaches to Linguistics often rely on the **introspective judgements** from just one or a few people
  - e.g. “is this sentence grammatical?”, “do these words mean the same thing?”
- Most linguists eventually have to confront the major insight of Sociolinguistics: **language varies extensively**
  - (Between people, communities, genres, conversation partners, etc.)
  - The same person might have different judgements at different times!
- This makes it questionable what the introspection of **any one person** can tell us about **language in general**
- Empirical techniques are inherently better suited for **generalizing from variable data**

# Ex: Categorical grammar rules

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate
  - We consider Kim an acceptable candidate

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate
  - We consider Kim an acceptable candidate
  - We consider Kim quite acceptable

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate
  - We consider Kim an acceptable candidate
  - We consider Kim quite acceptable
  - We consider Kim among the most acceptable candidates

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate
  - We consider Kim an acceptable candidate
  - We consider Kim quite acceptable
  - We consider Kim among the most acceptable candidates
  - *We consider Kim as an acceptable candidate

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate
  - We consider Kim an acceptable candidate
  - We consider Kim quite acceptable
  - We consider Kim among the most acceptable candidates
  - *We consider Kim as an acceptable candidate
  - *We consider Kim as quite acceptable

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate
  - We consider Kim an acceptable candidate
  - We consider Kim quite acceptable
  - We consider Kim among the most acceptable candidates
  - *We consider Kim as an acceptable candidate
  - *We consider Kim as quite acceptable
  - *We consider Kim as among the most acceptable candidates

# Ex: Categorical grammar rules

- A key technique of Generative syntax is to use **intuitive judgements** to figure out which sentences are **“grammatical” vs. “ungrammatical”**
- Examples from Pollard and Sag (1994)
  - We consider Kim to be an acceptable candidate
  - We consider Kim an acceptable candidate
  - We consider Kim quite acceptable
  - We consider Kim among the most acceptable candidates
  - *We consider Kim as an acceptable candidate
  - *We consider Kim as quite acceptable
  - *We consider Kim as among the most acceptable candidates
  - *?We consider Kim as being among the most acceptable candidates

# Ex: Categorical grammar rules

# Ex: Categorical grammar rules

- Actual examples from the *New York Times*:
  - The boys consider her as family and she participates in everything we do
  - I don’t consider it as something that gives me great concern
  - We consider that as part of the job
  - He considers them as having championship potential
  - Culturally, the Croats consider themselves as belonging to the West

# Ex: Categorical grammar rules

- Actual examples from the *New York Times*:
  - The boys consider her as family and she participates in everything we do
  - I don't consider it as something that gives me great concern
  - We consider that as part of the job
  - He considers them as having championship potential
  - Culturally, the Croats consider themselves as belonging to the West
- Investigating language from a **data-driven** approach reveals that it is **much more flexible** than the introspection-driven account

# Problems with categorical theories

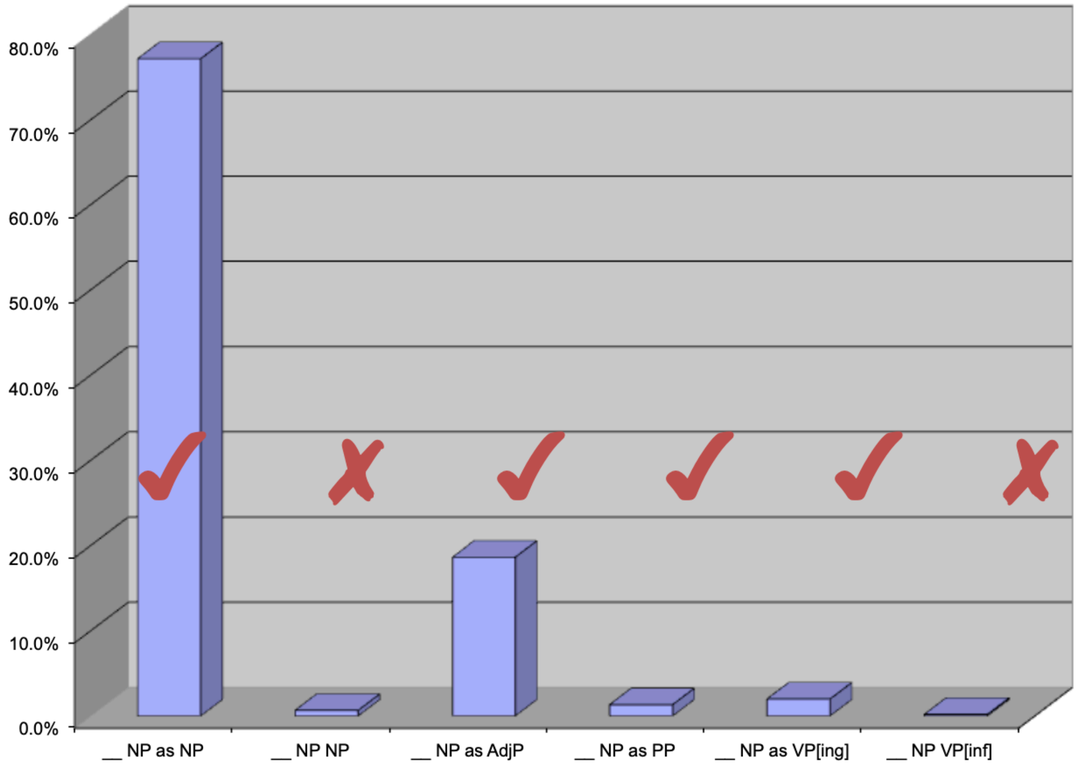
# Problems with categorical theories

- **They tend to claim too much**
  - They place a hard categorical boundary of grammaticality, where really there is a **fuzzy edge**, determined by many **conflicting constraints**

# Problems with categorical theories

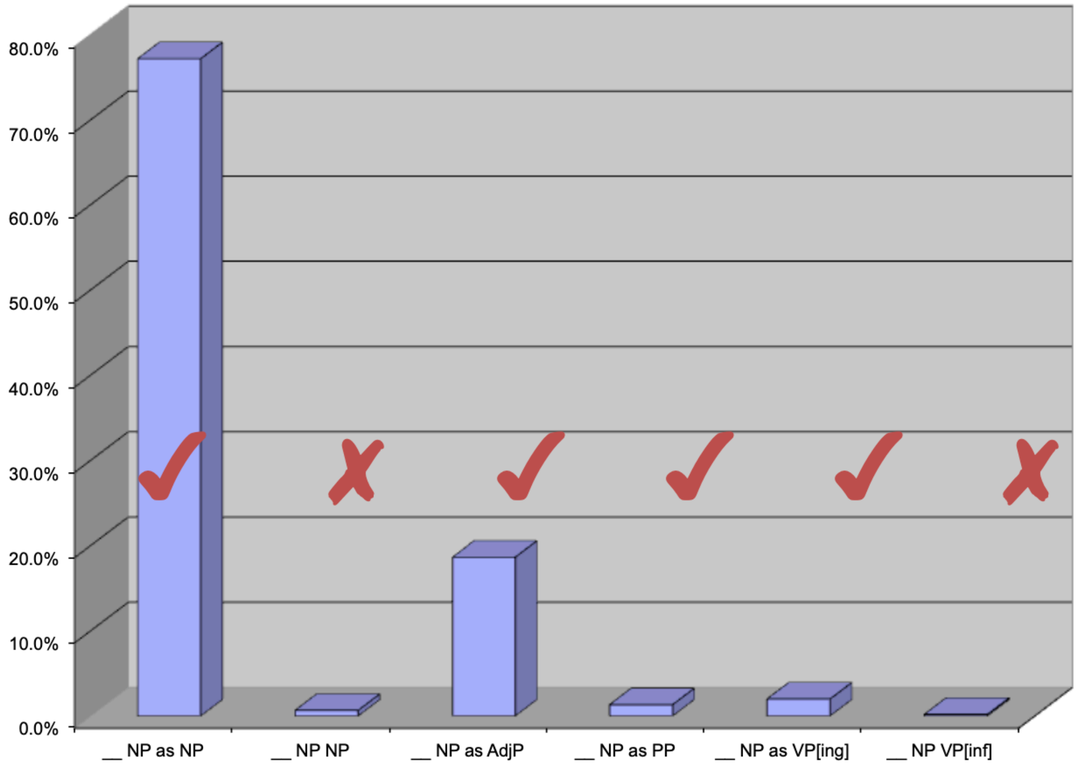
- **They tend to claim too much**
  - They place a hard categorical boundary of grammaticality, where really there is a **fuzzy edge**, determined by many **conflicting constraints**
- **They tend to explain too little**
  - They say very little about the **soft constraints** that explain how people **choose to say things** in given situations. These types of soft constraints have long been of interest to e.g. sociolinguists

# Probability distribution for *regard*



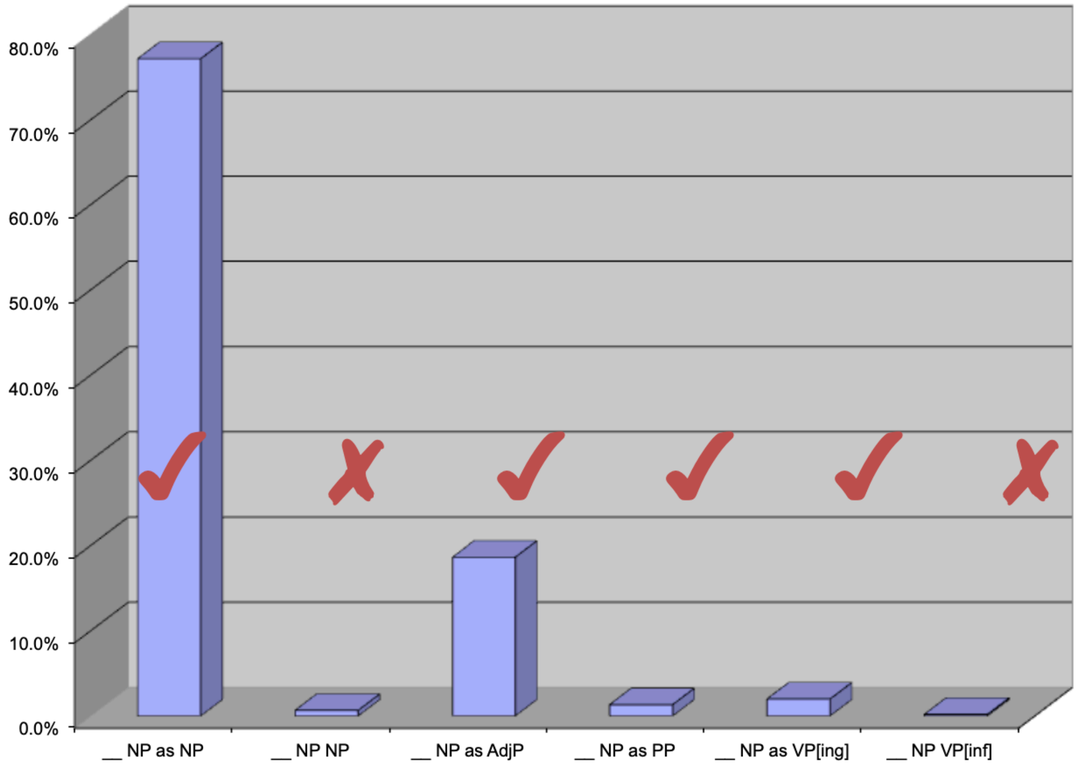
# Probability distribution for *regard*

- It might be more informative to construct a **probability distribution** over different variants of a sentence



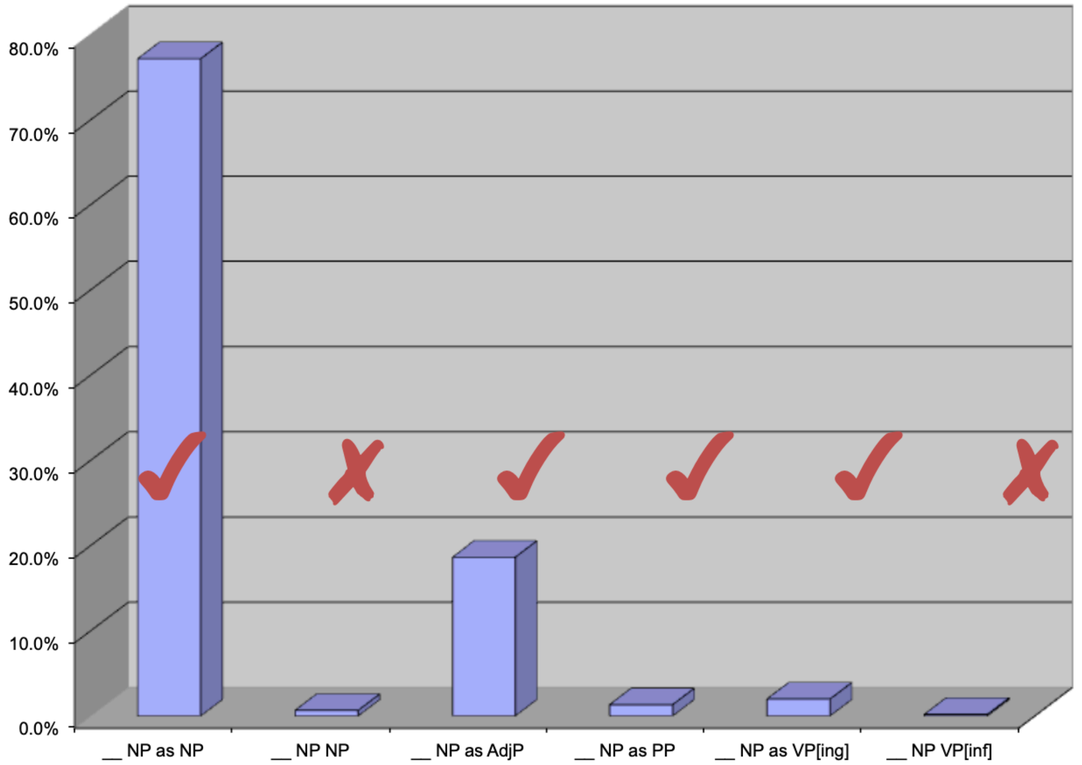
# Probability distribution for *regard*

- It might be more informative to construct a **probability distribution** over different variants of a sentence
- This distribution is calculated from a large **corpus** (dataset) of documented language “in the wild”



# Probability distribution for *regard*

- It might be more informative to construct a **probability distribution** over different variants of a sentence
- This distribution is calculated from a large **corpus** (dataset) of documented language “in the wild”
- We will discuss various standard corpora and **corpus linguistics methods** in this course



# About this course

# Course goals

# Course goals

- The overall theme of this course is **Language as Data**

# Course goals

- The overall theme of this course is **Language as Data**
  - How do we go from **raw text or audio** to **scientific insights?**

# Course goals

- The overall theme of this course is **Language as Data**
  - How do we go from **raw text or audio** to **scientific insights?**
  - What **levels of structure** exist in language data?

# Course goals

- The overall theme of this course is **Language as Data**
  - How do we go from **raw text or audio** to **scientific insights?**
  - What **levels of structure** exist in language data?
- We will introduce a toolkit to:

# Course goals

- The overall theme of this course is **Language as Data**
  - How do we go from **raw text or audio** to **scientific insights?**
  - What **levels of structure** exist in language data?
- We will introduce a toolkit to:
  - **Collect, clean, organize, and understand** linguistic data

# Course goals

- The overall theme of this course is **Language as Data**
  - How do we go from **raw text or audio** to **scientific insights?**
  - What **levels of structure** exist in language data?
- We will introduce a toolkit to:
  - **Collect, clean, organize, and understand** linguistic data
  - **Gain access** to existing corpora and understand **annotation conventions**

# Course goals

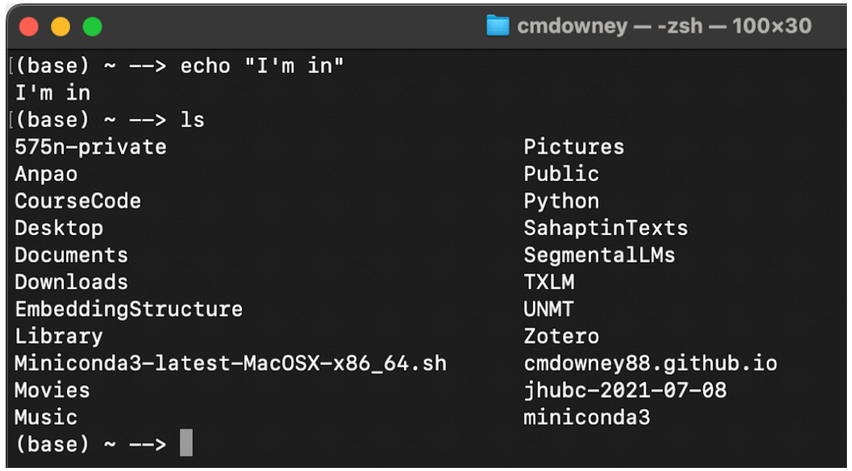
- The overall theme of this course is **Language as Data**
  - How do we go from **raw text or audio** to **scientific insights?**
  - What **levels of structure** exist in language data?
- We will introduce a toolkit to:
  - **Collect, clean, organize, and understand** linguistic data
  - **Gain access** to existing corpora and understand **annotation conventions**
  - **Visualize** patterns in language, and **effectively communicate** your findings

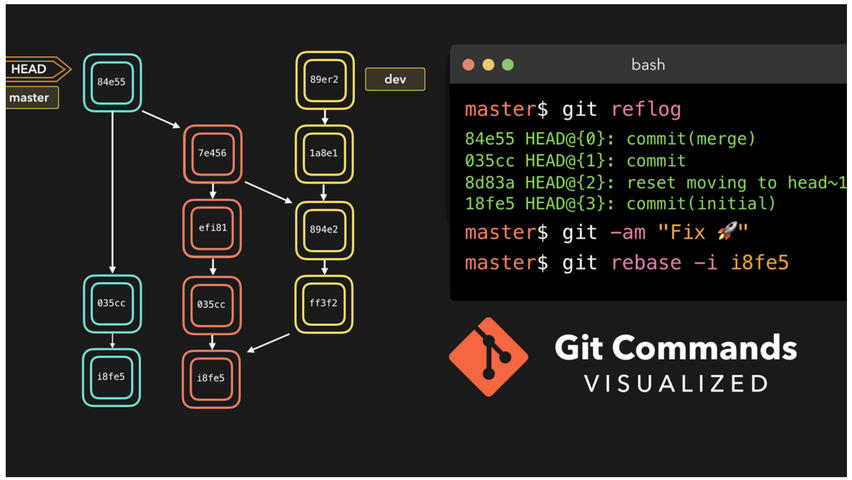
# Course goals

- The overall theme of this course is **Language as Data**
  - How do we go from **raw text or audio** to **scientific insights?**
  - What **levels of structure** exist in language data?
- We will introduce a toolkit to:
  - **Collect, clean, organize, and understand** linguistic data
  - **Gain access** to existing corpora and understand **annotation conventions**
  - **Visualize** patterns in language, and **effectively communicate** your findings
  - **Test linguistic hypotheses** with statistical robustness

# Larger theme: Core Data Science Tools

- Covered in the first part of the course
- Using the **command line**
  - Will learn how to interface with computers in a new and **more powerful** way
  - Is sometimes the **only way** to interface with powerful servers for data science!
- **Version control** with **Git and Github**
  - The standard way to manage **collaboration, edits,** and **differing versions** of software



# Larger theme: Data manipulation in Python

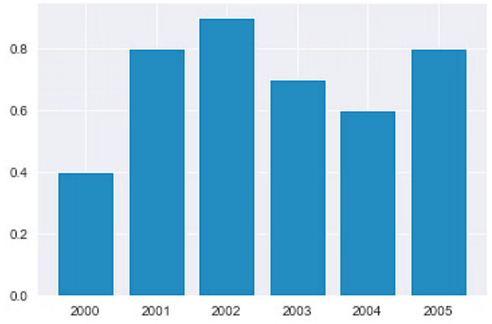
- Very basic introduction to **Python**
  - The dominant language for **Science** and **Data Engineering**
  - Easy language to **get started**
- Introduction to relevant **Python libraries**
  - Regular Expressions
  - Natural Language Tool Kit
  - Pandas

```
>>> nltk.corpus.sinica_treebank.tagged_words()
[('ä', 'Neu'), ('åæ', 'Nad'), ('åç', 'Nba'), ...]
>>> nltk.corpus.indian.tagged_words()
[('মহিষের', 'NN'), ('সন্তান', 'NN'), (':', 'SYM'), ...]
>>> nltk.corpus.mac_morpho.tagged_words()
[('Jersei', 'N'), ('atinge', 'V'), ('m\xe9dia', 'N'), ...]
>>> nltk.corpus.conll2002.tagged_words()
[('Sao', 'NC'), ('Paulo', 'VMI'), ('(', 'Fpa'), ...]
>>> nltk.corpus.cess_cat.tagged_words()
[('El', 'da0ms0'), ('Tribunal_Suprem', 'np0000o'), ...]
```

```
years = range(2000, 2006)
apples = [0.35, 0.6, 0.9, 0.8, 0.65, 0.8]
oranges = [0.4, 0.8, 0.9, 0.7, 0.6, 0.8]
```

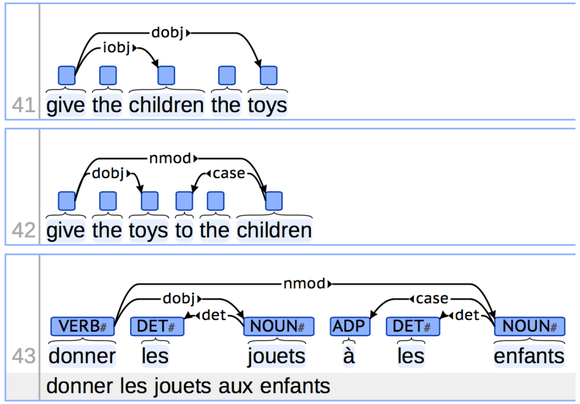
```
plt.bar(years, oranges)

plt.xlabel('Year')
plt.ylabel('Yield (tons per hectare)')

plt.title("Crop Yields in Kanto")
```

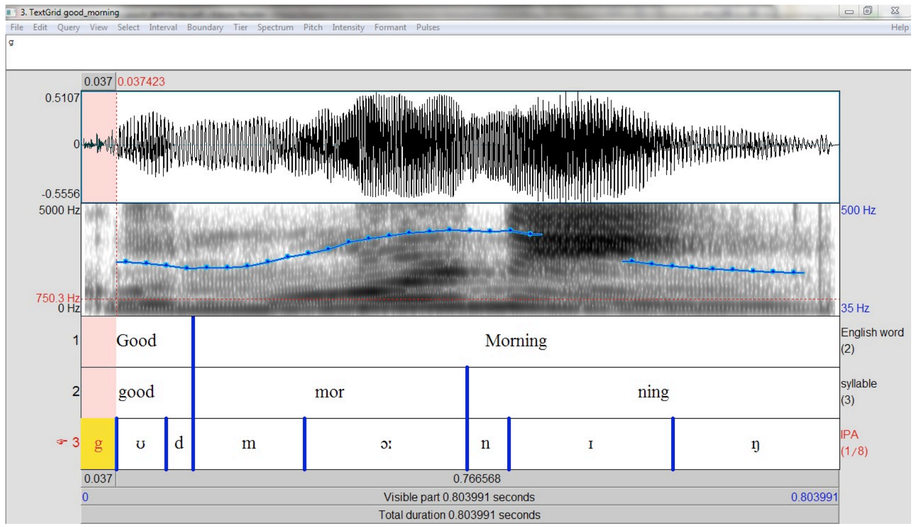
```
<BarContainer object of 6 artists>
```



# Larger theme: Linguistic annotation

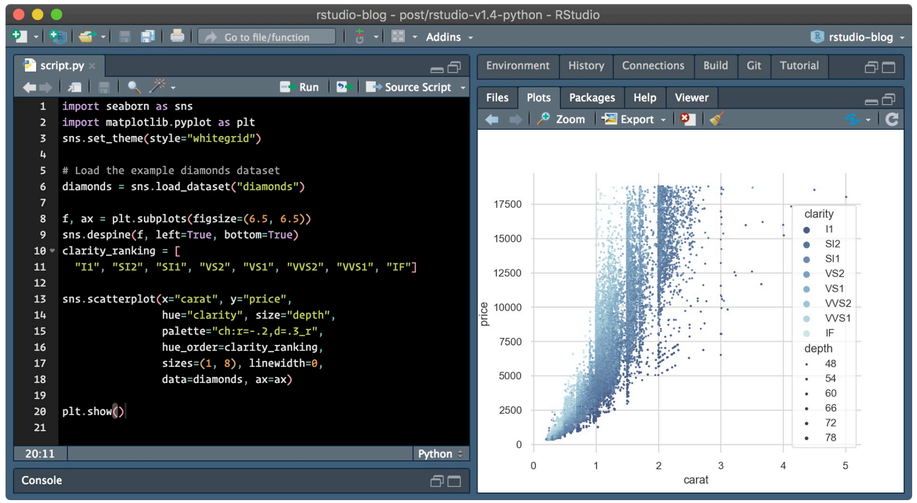
- Introduction to **common annotation schemes** for Linguistic data
- Discussion of why it is important to have **machine-readable** data
- Discussions of the **challenges** and **reliability** of annotation

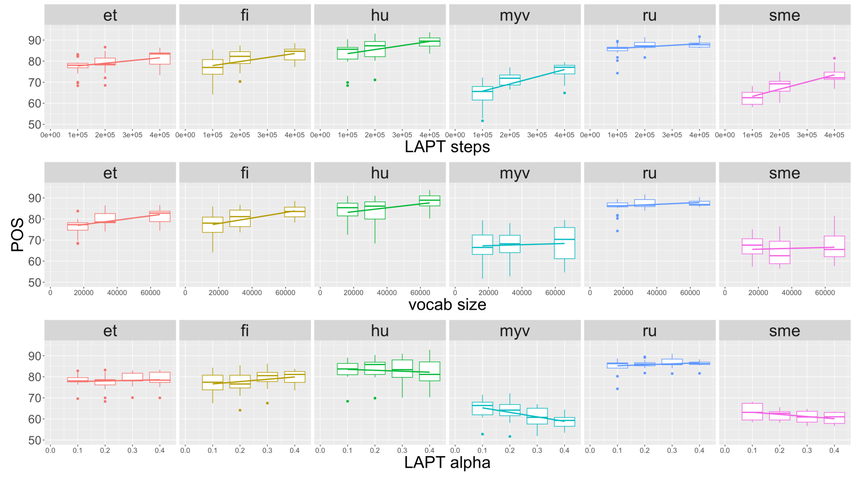


# Larger theme: R for testing and visualization

- Introduction to **R**
  - Most commonly used language for **statistical testing**
  - Also has the best libraries for **data visualization**
- **ggplot**
  - Library for developing **professional data visualizations**
  - Tricky at first, but gives the best results



# Policies and Organization

# Basic Information

- Instructor
  - C.M. Downey
  - Assistant Professor in Linguistics & Data Science
- Meeting time
  - Monday/Wednesday 9:00-10:15am
  - **This week only** we meet **Wednesday/Friday**
- **Office hours**: by appointment
  - Feel free to reach out to schedule an appointment any time!

# Online Resources

# Online Resources

- **Blackboard**
  - Announcements
  - Discussion boards
  - Homework submission

# Online Resources

- **Blackboard**
  - Announcements
  - Discussion boards
  - Homework submission
- **Course website** (cmdowney88.github.io/teaching/ling250/spring2026)
  - Detailed course schedule
  - Syllabus
  - Slides (posted after class)
  - Important links and documents

# “Required” Textbooks

# “Required” Textbooks

- You will sometimes be assigned **required readings** before class, but these will always be drawn from **free online sources**
  - Available online as PDFs

# “Required” Textbooks

- You will sometimes be assigned **required readings** before class, but these will always be drawn from **free online sources**
  - Available online as PDFs
- **Links** to these sources can be found here and on the **course website:**
  - *Natural Language Processing with Python*
  - *Data Science from Scratch*
  - *Speech and Language Processing*

# Assessment Structure

# Assessment Structure

- **40% Term Project**
  - Substantial research project culminating in a paper (more later)

# Assessment Structure

- **40% Term Project**
  - Substantial research project culminating in a paper (more later)
- **20% Midterm Project**
  - Structured data-processing task, meeting a precise output specification

# Assessment Structure

- **40% Term Project**
  - Substantial research project culminating in a paper (more later)
- **20% Midterm Project**
  - Structured data-processing task, meeting a precise output specification
- **20% Homework**
  - Small written and programming exercises

# Assessment Structure

- **40% Term Project**
  - Substantial research project culminating in a paper (more later)
- **20% Midterm Project**
  - Structured data-processing task, meeting a precise output specification
- **20% Homework**
  - Small written and programming exercises
- **20% Participation**
  - 10% Attendance
  - 10% Engagement in Discussion/Activities

# Term Project

# Term Project

- Completed **alone** or in a **small group (2-3 students)**

# Term Project

- Completed **alone** or in a **small group (2-3 students)**
- Conduct a topical **research project**, which must
  - **Apply** course techniques to a problem with **language as data** (pick something you're genuinely interested in!)
  - Pose and test a **scientific hypothesis**
  - Culminate in a **research paper** and **presentation**

# Term Project

- Completed **alone** or in a **small group (2-3 students)**
- Conduct a topical **research project**, which must
  - **Apply** course techniques to a problem with **language as data** (pick something you're genuinely interested in!)
  - Pose and test a **scientific hypothesis**
  - Culminate in a **research paper** and **presentation**
- **Deliverables**: writeup, presentation, code repository
  - Scaffolded with **incremental milestones** due throughout the semester
  - First step: an **interest survey** due **next Thursday (1/29)**

# Attendance/Participation

# Attendance/Participation

- I will keep track of attendance, which will be **visible on Blackboard**

# Attendance/Participation

- I will keep track of attendance, which will be **visible on Blackboard**
- You have **two "free" absences** - for any reason, no questions asked
  - **No need to notify me**. I'll automatically account for these at the end of the term

# Attendance/Participation

- I will keep track of attendance, which will be **visible on Blackboard**
- You have **two "free" absences** - for any reason, no questions asked
  - **No need to notify me**. I'll automatically account for these at the end of the term
- I'll also make exceptions for **important obligations/circumstances**
  - e.g. civic, religious, or family obligations; illness; job interviews; conferences

# Attendance/Participation

- I will keep track of attendance, which will be **visible on Blackboard**
- You have **two "free" absences** - for any reason, no questions asked
  - **No need to notify me**. I'll automatically account for these at the end of the term
- I'll also make exceptions for **important obligations/circumstances**
  - e.g. civic, religious, or family obligations; illness; job interviews; conferences
- Any other absences will **count against your 10% attendance grade**

# Attendance/Participation

- I will keep track of attendance, which will be **visible on Blackboard**
- You have **two "free" absences** - for any reason, no questions asked
  - **No need to notify me**. I'll automatically account for these at the end of the term
- I'll also make exceptions for **important obligations/circumstances**
  - e.g. civic, religious, or family obligations; illness; job interviews; conferences
- Any other absences will **count against your 10% attendance grade**
- **10%** of grade for **engagement in discussions / check-in activities**

# Participation

# Participation

- Participation is also **part of your final grade** (10%)

# Participation

- Participation is also **part of your final grade** (10%)
- A large part of class time will be devoted to **hands-on demonstrations** of various software tools
  - You are expected to **follow along** with the demo on your **own device**
  - This means you will need to bring a **laptop** or **other appropriate device** to class (a tablet might work if it is running Windows; an iPad probably won't work)

# Participation

- Participation is also **part of your final grade** (10%)
- A large part of class time will be devoted to **hands-on demonstrations** of various software tools
  - You are expected to **follow along** with the demo on your **own device**
  - This means you will need to bring a **laptop** or **other appropriate device** to class (a tablet might work if it is running Windows; an iPad probably won't work)
- **Before Friday:** follow the **setup guide** posted on Blackboard and the course website to get your device ready

# Academic Honesty

# Academic Honesty

- All **standard university policies** apply

# Academic Honesty

- All **standard university policies** apply
- **IMPORTANT: Generative AI Policy**
  - **NOT allowed for***: homework, midterm project, project milestones, project writeup
  - **allowed for**: programming work on the **term project only**
  - Rule of thumb: using AI to **learn and clarify concepts** is fine; using it to **generate work you submit as your own** is **not**
    - ***You can use AI to **help you learn** on assignments where it's not otherwise allowed; just don't have it do the work for you
  - **Honor system** - I don't want to be the AI police

# Pre-requisites and Tools

# Pre-requisites and Tools

- What I assume:
  - **Some familiarity** with **Python Programming**
  - A **laptop or other device** on which you can conduct computational work

# Pre-requisites and Tools

- What I assume:
  - **Some familiarity** with **Python Programming**
  - A **laptop or other device** on which you can conduct computational work
- Computing resources
  - I will provide **access to UR's supercomputing cluster (BlueHive)**
  - Part of this class is learning how to use **command-line interfaces**
    - This is still important for **industry and research jobs**
  - Not mandatory to use, but helpful for intensive computing

# Deadlines and Late Work

# Deadlines and Late Work

- Unless specified otherwise, all assignments are due **11pm Eastern**

# Deadlines and Late Work

- Unless specified otherwise, all assignments are due **11pm Eastern**
- Work submitted after the deadline will incur a penalty
  - Up to **1 hour** late: -5%
  - Up to **24 hours** late: -10%
  - Up to **48 hours** late: -20%
  - **> 48 hours** late: not graded (0 for the assignment)

# Deadlines and Late Work

- Unless specified otherwise, all assignments are due **11pm Eastern**
- Work submitted after the deadline will incur a penalty
  - Up to **1 hour** late: -5%
  - Up to **24 hours** late: -10%
  - Up to **48 hours** late: -20%
  - **> 48 hours** late: not graded (0 for the assignment)
- Please **feel free to request extensions**, but you must request it **before the deadline**
  - I will be more willing to grant extensions the longer before the deadline it is requested

# Questions/Discussion?