


Introduction to Data Science with R

BSDS 100, Spring 2021
Michael Ruddy

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

About me

- Michael Ruddy (he/him)
- PhD in Mathematics in 2019 from North Carolina State University
- Currently in Minneapolis, Minnesota
- Research in Geometry and Data Science
- Bouldering, Gardening, Cooking

Data: What? Why?

- Data: “information gathered through observation”
- Rarely ever the full picture
 - Sampling
 - Future events
 - Errors in observation
- Understand the unknown from this restricted view

Conclusions from Data

Inference

- Understand relationship between variables
- Understand population from sample observations
- Statistics
 - Machine Learning

Prediction

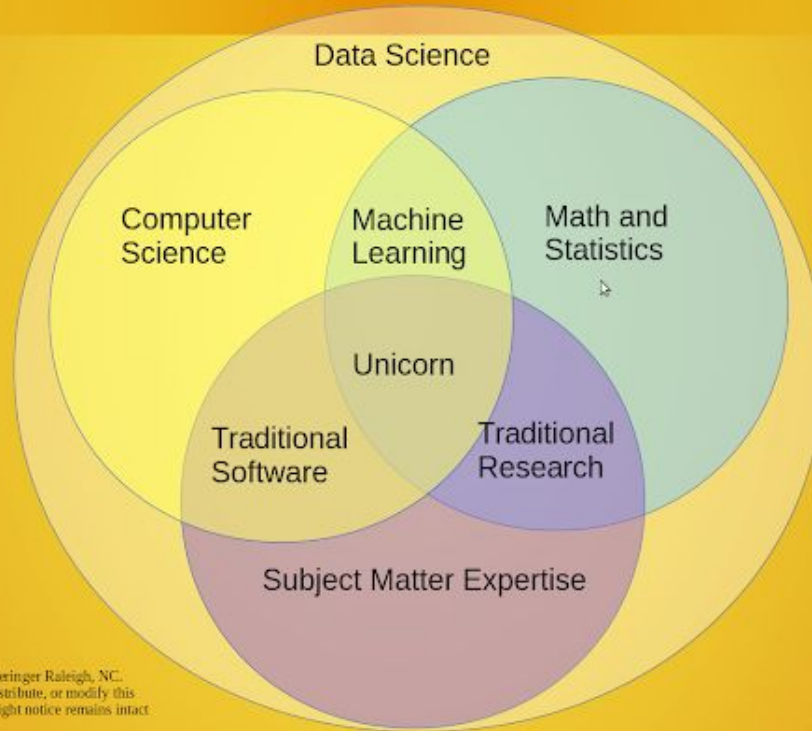
- Predict outcomes from new data points
- Understand uncertainty in our Prediction
- Machine Learning
 - Statistics

Why now?

- Statistics says hello
- We can collect and store massive amounts of data
- We have much more computer power
- Artificial Intelligence

Data Science: What? Who?

Data Science Venn Diagram v2.0



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this
image, provided that this copyright notice remains intact.

Data Science: What? Who?

- “Fields of research are defined by the people who participate”
- Things Data Scientists sometimes do:
 - Gather data
 - Clean data
 - Manage data storage
 - Exploratory analysis of data
 - Make predictions/inferences from data
 - Summarize/Communicate/Visualize data

Ethics in Data Science

- Center for Applied Data Ethics at USF
- Ethical collection of sensitive data
- Your predictions/inferences are only as good as your data!
- Unintended consequences of predictions from data
 - Can reinforce structural inequities and systemic racism

Ethics in Data Science

- Center for Applied Data Ethics at USF
- Ethical collection of sensitive data
- Your predictions/inferences are only as good as your data!
- Unintended consequences of predictions from data
 - Can reinforce structural inequities and systemic racism
- Example: Predictive Policing
 - Predict time/location of crime from arrest statistics and demographic data
 - Arrest stats are an imperfect proxy for crime stats
 - Feedback loop confirms bias in data

This course

- Learn how to use R programming language to perform different aspects of 'data science' to varying degrees
- Lots of statistical and graphical tools
- Free, open source, actively-maintained
- Great GUIs: RStudio, Jupyter Notebook
- Why not Python?

This course

- No programming or statistical background is needed
- We will learn a small fraction what is possible in R and using tools from data science.
- More resources:
 - Flowingdata
 - Coursera
 - Kaggle

Questions about programming in R?

- Google

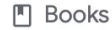
Questions about programming in R?

- Google

```
IndexError                                Traceback (most recent call last)
<ipython-input-3-e4d12b5190dc> in <module>
    12     for file in os.listdir(str(mask_path)):
    13         split_file = file.split('_')
--> 14         if (split_file[0] == patient_no) & (split_file[1] in masks):
    15             patient_masks.add(split_file[1])
    16     if patient_masks == set(masks):

IndexError: list index out of range
```

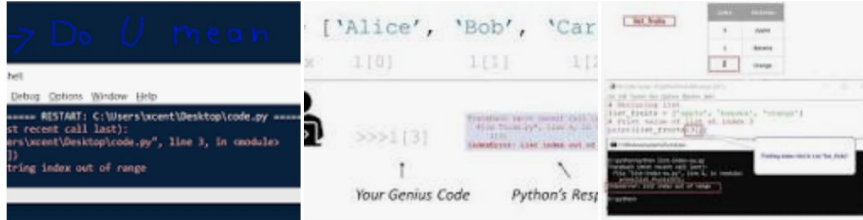
IndexError: list index out of range



Settings

Tools

About 551,000 results (0.61 seconds)



View all

The error “**list index out of range**” arises if you access invalid indices in your Python **list**. For example, if you try to access the **list** element with **index** 100 but your lists consist only of three elements, Python will throw an **IndexError** telling you that the **list index** is **out of range**.

blog.finxter.com › python-indexerror-list-index-out-of-ra...

Python IndexError: List Index Out of Range (How to Fix This ...)



About featured snippets



Feedback

stackoverflow.com › questions › indexerror-list-index-... ▼

IndexError: list index out of range and python - Stack Overflow

Questions about programming in R?

- Google
- Office Hours
- Canvas Discussion board

Using Google and other resources

- Do ***NOT*** copy code without crediting the original source
- In this course, do not copy/paste ***any*** of your classmates' code
- Do not replicate code without understanding it
- Disclose all sources of outside help on assignments (classmates, StackExchange posts, etc.)