
Python for Data Analysis

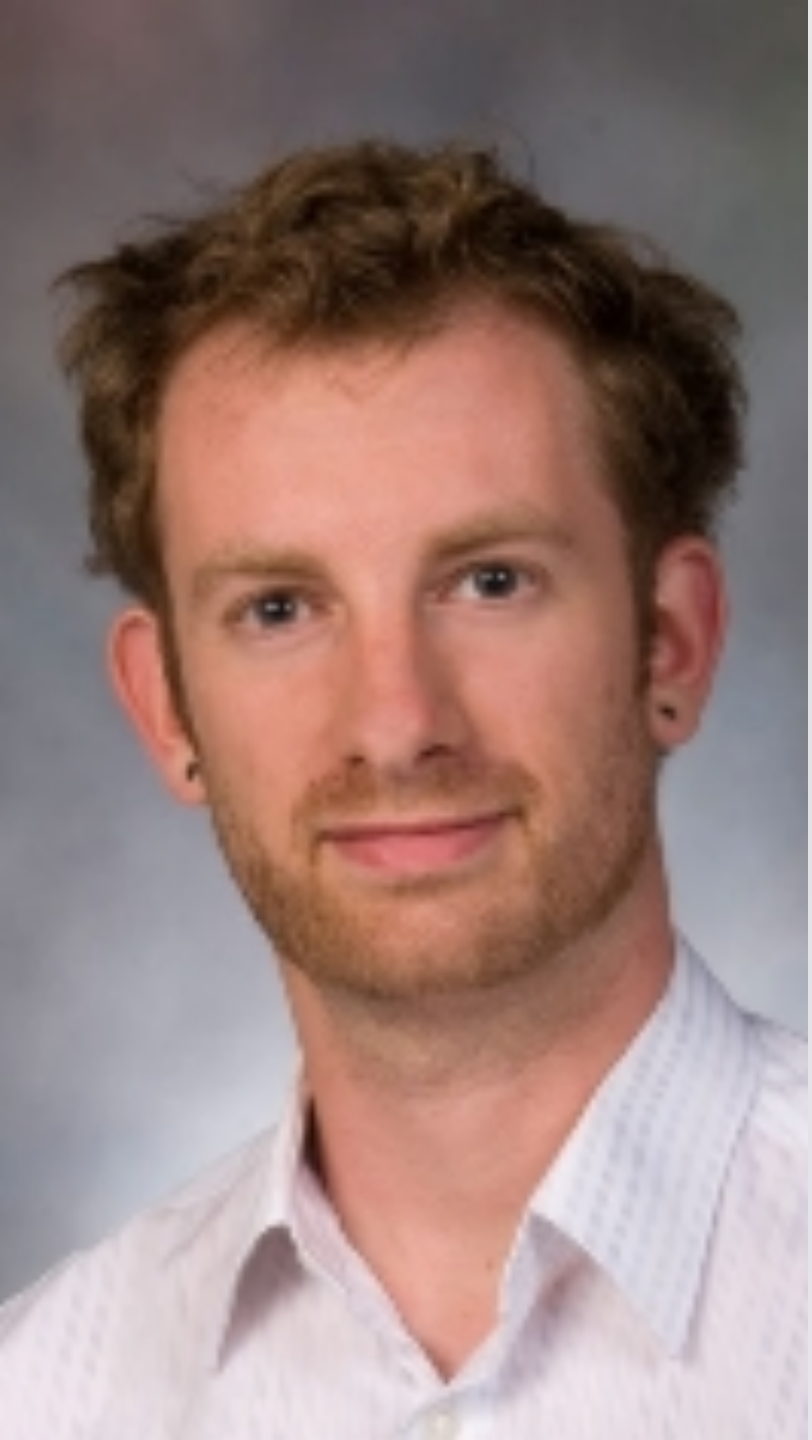
# Introduction

- Learned MATLAB in college
- Started out as a business analyst
- Learned R to apply random forests
- Listened to a lecture by Hadley Wickham
- …
- Invited conference speaker
- Working as a data scientist at Slalom

# Agenda

- What is data analysis?
- Why learn to program?
- Why learn to program in python?
- Code Walkthrough & Exercise
- Resources

# What is
# data analysis?



*Data analysis is the process by which data becomes understanding, knowledge and insight.*

Hadley Wickham

1. Define
2. Collect
3. Transform
4. Explore
5. Model
6. Communicate

# Why learn to program?

# #1

# Provides liberation to analyze data on your own terms

# #2

# Provokes knowledge sharing and reproducibility

# #3

# Promotes continuous improvement through automation

# Why program in Python?

# #1

# Python is free

# #2

# Python is a general purpose language

# #3

# Python is designed for ease of use

```
>>> import this
The Zen of Python, by Tim Peters

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!
```

# #4

# Python has an organized development community

# Python has a rich and cohesive package ecosystem for conducting data science

## Analysis

**NumPy**
[Arrays]

**Pandas**
[Data Frames]

## Extraction

**Beautiful Soup**
[Web-Scraping]

**NLTK**
[NLP]

## Modeling

**Scikit-learn**
[Machine Learning]

**StatsModels**
[Statistics]

# Code Walkthrough & Exercises

# Stay in touch

- Twitter: @josiahjdavis
- E-mail: josiah.j.davis@gmail.com
- Website: www.josiahjdavis.com