PHS Data Science Workshop 2022

Day 1

About us

[University of Utah Department of Population Health Sciences]

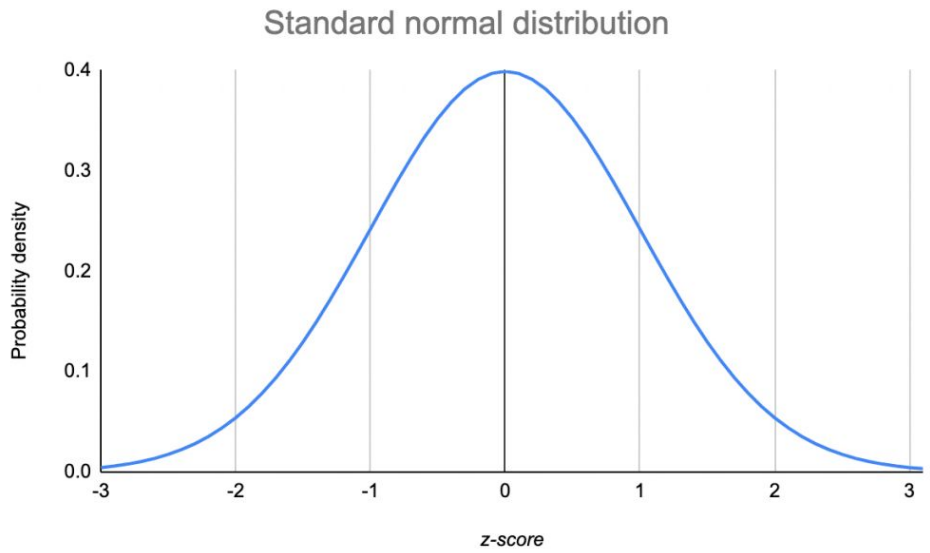
Instructor: Alec Chapman

Teaching Assistant: Christian Dalton

Administrative Support: Marcie Leek + Josh Taylor

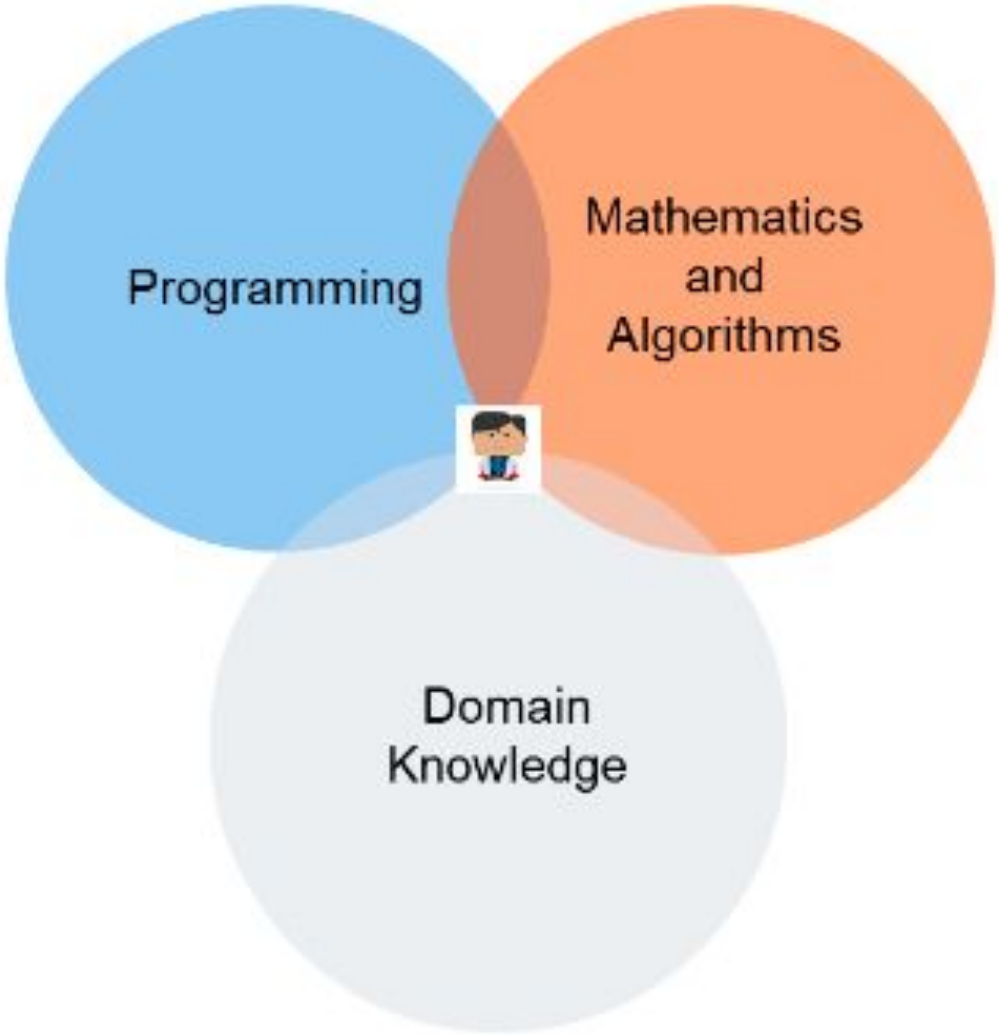
Faculty Advisor: Daniel Scharfstein

1. What is data science?



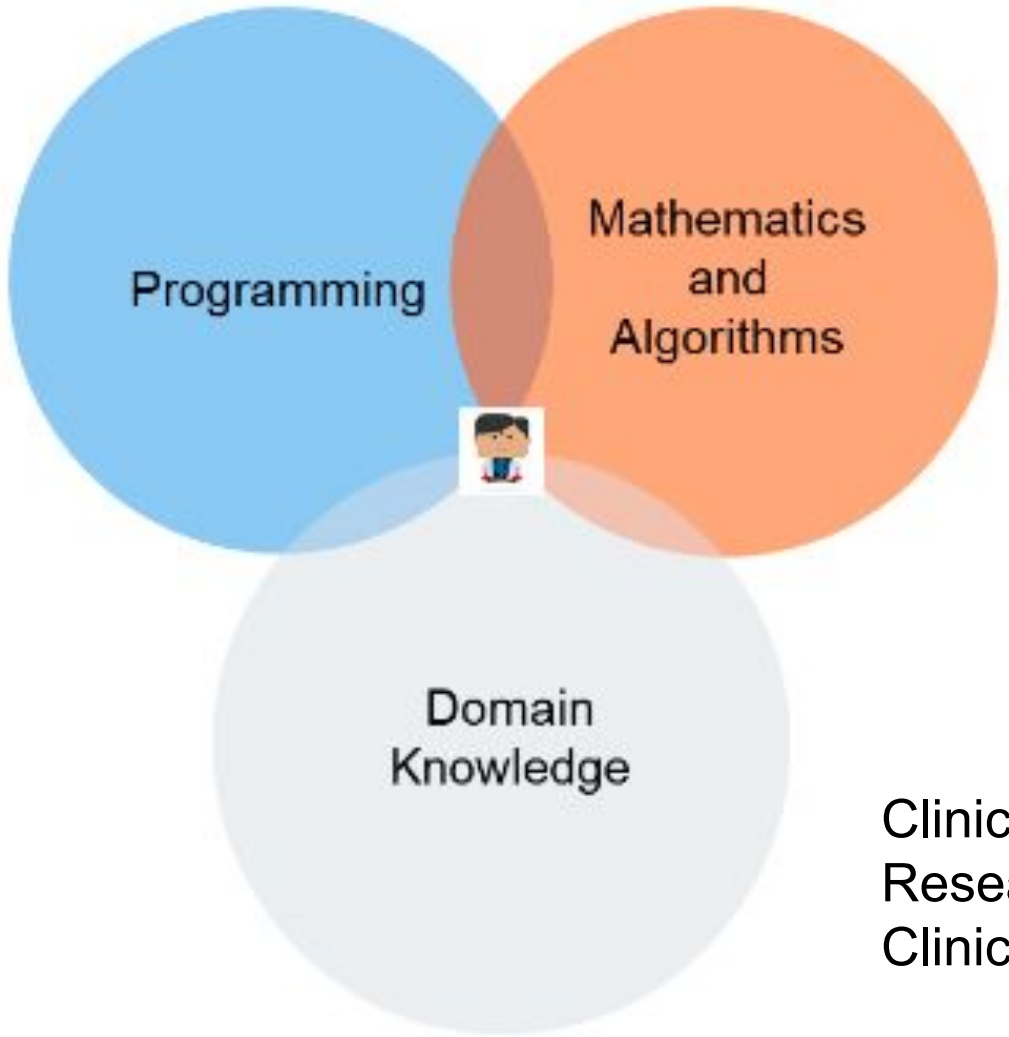
Our working definition of data science will be:

"A process of obtaining, transforming, and analyzing data to understand the world."



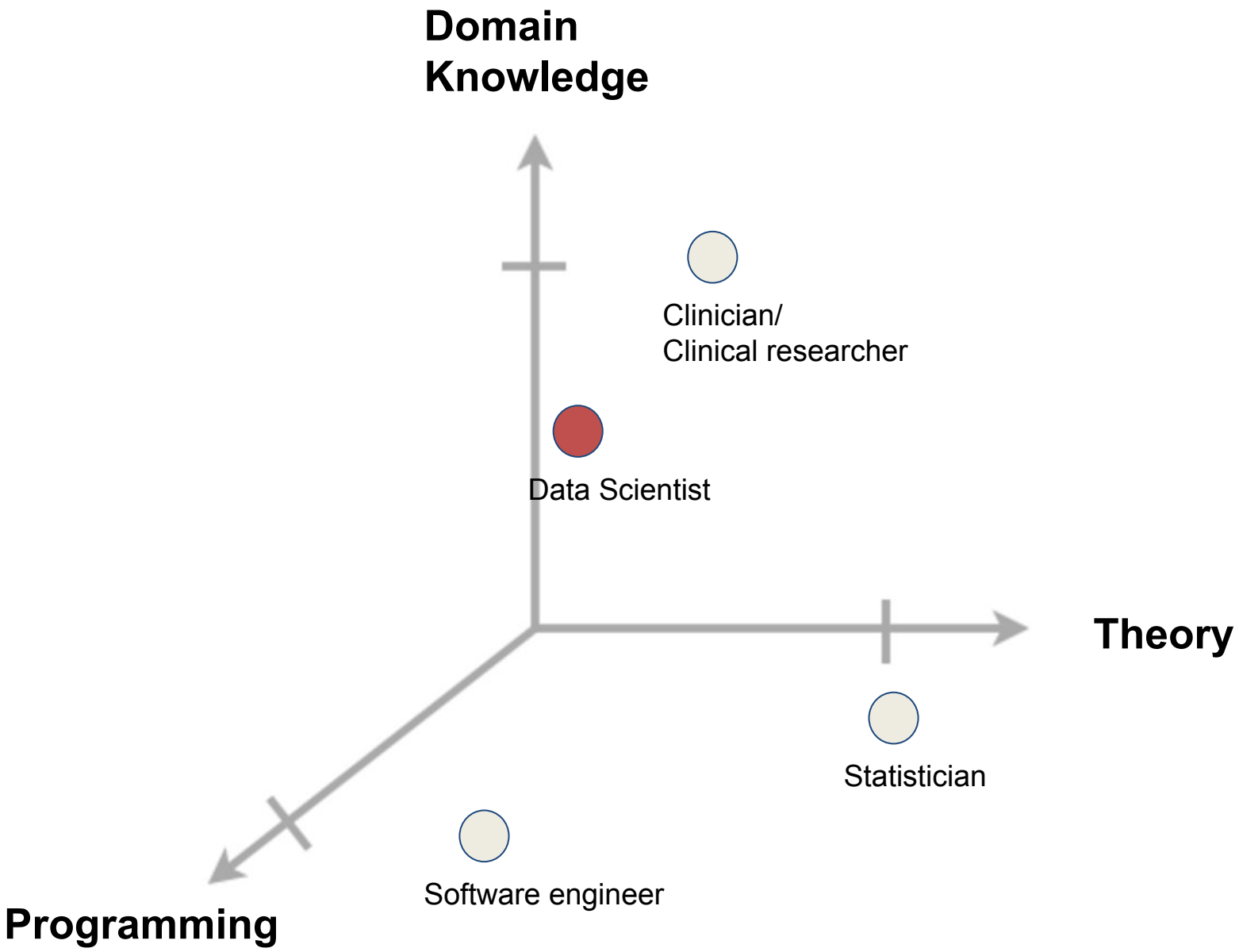
Software development
Databases
R, SQL, Python

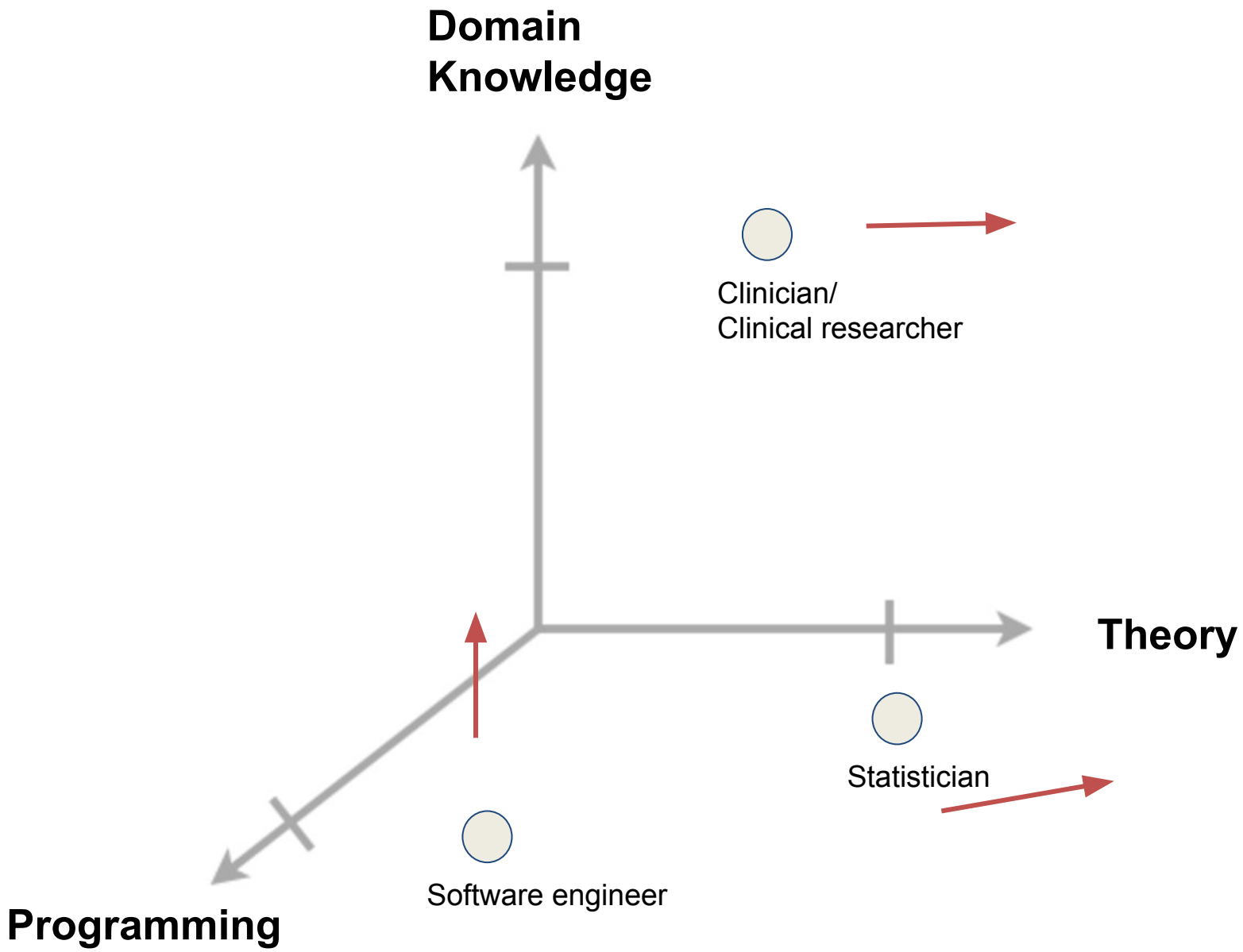
Statistics
Theory



Clinical workflows
Research questions
Clinical data

2. Why learn data science?



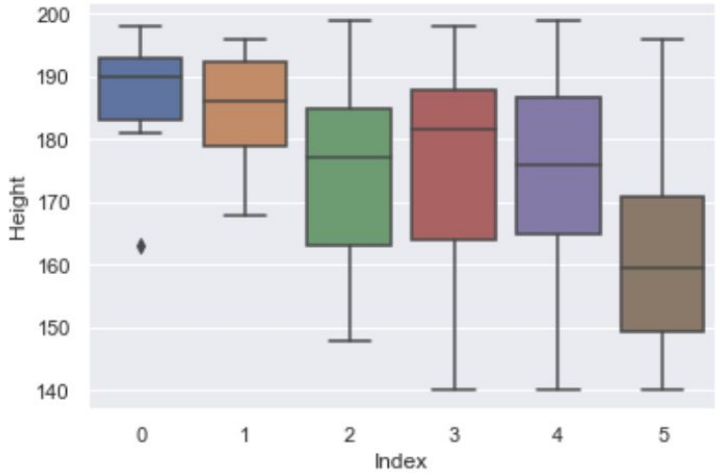


3. How to learn data science

By *doing*

We could also plot individual pairs using `seaborn`:

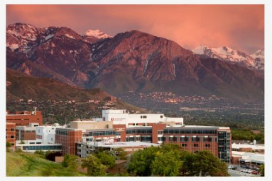
```
In [42]: sns.boxplot(x="Index", y="Height", data=df)
Out[42]: <AxesSubplot:xlabel='Index', ylabel='Height'>
```



jupyter 00-Roadmap-Module-1 Last Checkpoint: 08/06/2022 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Markdown



University of Utah
Population Health Sciences
Data Science Workshop

HEALTH
UNIVERSITY OF UTAH

Module 1

Welcome to the PHS Data Science Workshop! This class will use **Jupyter Notebooks** for writing code, running it on data, and viewing the results. Each day will start with a **Roadmap** that will outline the plans for the day.

The general outline for each day will be:

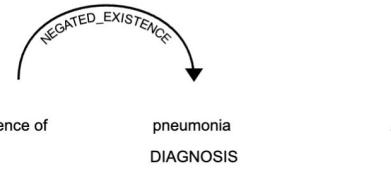
1. Introduction of today's topic and review what we learned the day before
2. Learn the skills of the day by completing educational notebooks
3. Apply our new skills to a use case of pneumonia

```
In [90]: from medspacy.visualization import visualize_dep, visualize_ent
In [91]: visualize_ent(doc)
```

There is no evidence of NEGATED_EXISTENCE pneumonia DIAGNOSIS .

The second, `visualize_dep` one visualizes the relationships between targets and modifiers in a dependency-style visualization. Here, we can see that the modifier **"no evidence of"** is applied to the target **"pneumonia"**.

```
In [92]: visualize_dep(doc)
```



Roadmap

Day 1. Python Essentials

Day 2. Clinical Databases and SQL

Day 3. NLP + Machine Learning

What we'll do in this class

The *least* fun stuff…

- Installation
- Setting up the environment
- Troubleshooting

The *more* fun stuff…

- Learning Python syntax
- Data structures
- Exploring libraries

The *really* fun stuff

- Exploring clinical data
-

Day 1

- Python Syntax
- Data types and structures
- Writing functions
- Using important Python libraries

Day 1.
Python
Essentials

Day 2

- Relational databases
- SQL syntax
- Transforming and aggregating data

Day 2. Clinical Databases and SQL

Day 3

- Learn how to read clinical notes
- Develop an NLP system to extract information from notes
- Build a complete NLP system for identifying pneumonia in radiology reports

- Train a machine learning model to predict diabetes

Day 3.
NLP + Machine Learning