

# A Hands-on Introduction to End-to-end Data Projects in Python for SAS Programmers

---

Isaiah Lankham

Legacy Health  
System Office  
Portland, OR

Matthew T. Slaughter

Kaiser Permanente  
Center for Health Research  
Portland, OR



**SEIZE THE DATA**  
**SACRAMENTO, CA**

HOW-170 • 4:00 p.m. • 05 September 2024

**WUSS**

WESTERN USERS OF SAS SOFTWARE

SEPTEMBER 4TH - 6TH 2024 | [WWW.WUSS.ORG](http://WWW.WUSS.ORG)

# There's a reason it's a meme...

Lots of googling for  
the right Stack  
Overflow posts

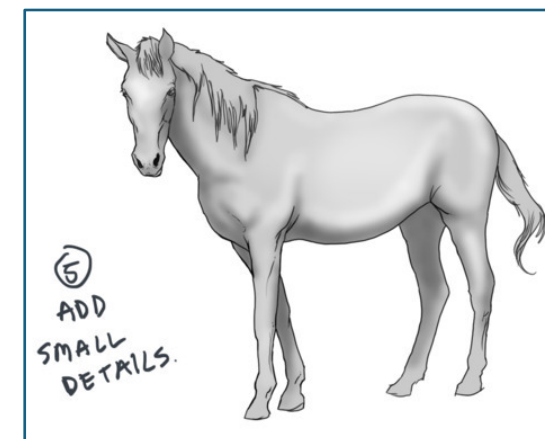
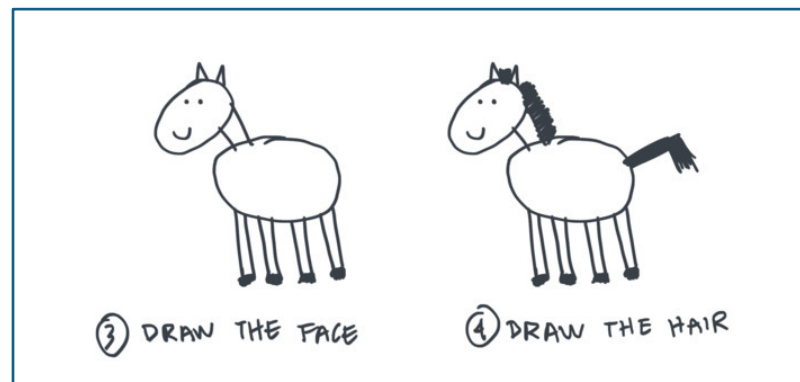
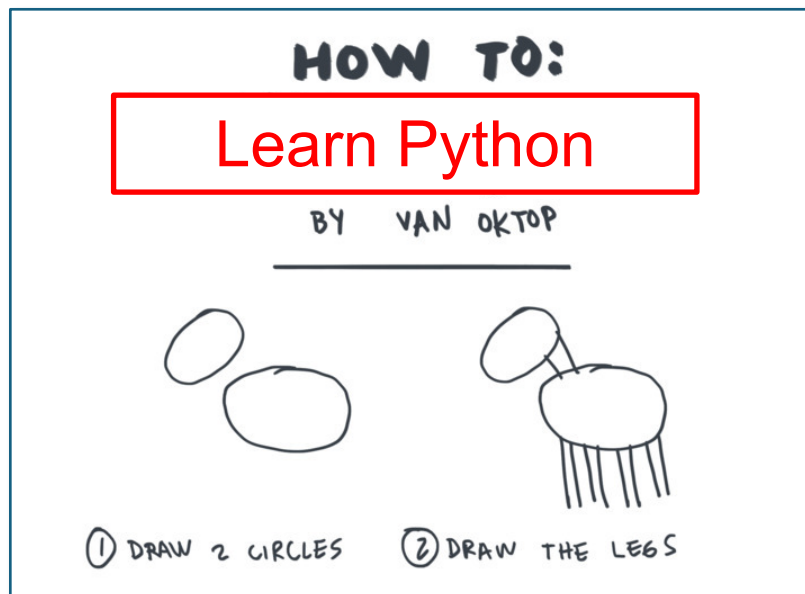


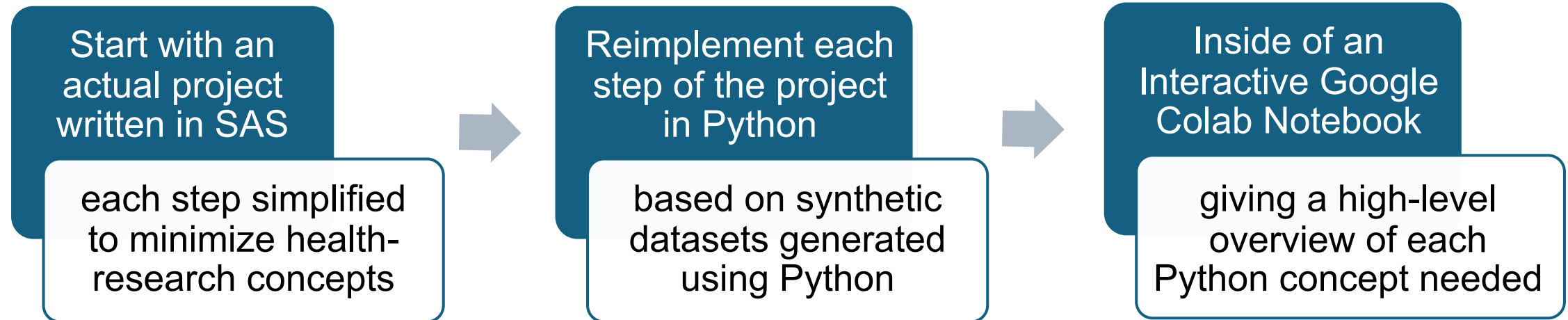
image source: <https://oktop.tumblr.com/image/15352780846>

# Congrats!

---

**Scenario:** It's your first day as an intern for the Research Center

**Today's Agenda:** Complete an end-to-end data project in Python



# Project Steps Overview

---

## Section 1. Data Access

- Load .csv files from GitHub for patients, medication dispenses, and diagnoses (dx)

## Section 2. Data Exploration

- Check each column in each dataset for anomalous values

## Section 3. Data Cleaning

- Correct anomalous values

## Section 4. Data Munging

- Create a single analytic dataset with columns from each "cleaned" dataset

## Section 5. Data Management

- Save the analytic dataset to disk in multiple formats

## Section 6. Data Analysis

- Build a logistic regression model

# Google Colab Overview

