Tidy Data format:

* Rows- subjects
* Columns- Variable
* Sample space
* Every row is a subject and every column is a specific variable (no empty rows, no merged columns)

GitHub (?) (bought by microsoft)

* Used to upload data/code/image/spreadsheet files
* Repo (repository) - similar to like a folder in google drive :)
  + Can make private repo (but microsoft will still have access/be able to see)
* GitHub Copilot (AI used to create/ write code)
  + Revolutionized data science

Visual StudioCode (VS Code)

IPYNB (something python notebook- it’s the code)

Boxes are “cells”

Python

Java

Visual studio code

Google Collab/ Google notebook (has code !!)

Html - type of computer language

Markup language (#)

* Simplified version of HTML

Code Package/Library

API- Application … Interface

% is called a magic command

Review chi square test

Jupiter Notebooks incorporate normal english 😀 and can be shared with others (what professor uses in class)

Restricted/Open data

Red Cap

Look into working for health insurance ??

Anonymized data is open data

Find youtube video on deanonymization of data

EHR- Electronic Health Records

Can access the IRB portal for undergraduate student research

**HOMEWORK:**

* Look at Behavioral Risk Factor Surveillance System
* National Health and nutrition Examination Survey
* Health Survey for England

GPU- graphics processing units (NVidia stock prices) (AMD)

* Used to be for games, now used for data

Pandas

Wbgapi

data.worldbank.org