



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Biostatistics BT2023

Lecture 2 : UNIX programming

Himanshu Joshi 05 August 2022



Statistics

Descriptive statistics

Summarize important characteristics of a set of measurements

Inferential statistics

Procedure for making inferences about the population's characteristics drawn from the sample



Population, Sampling and statistical inference



Sampling has to random

Example: Poll results by calling few homes

Average salary by collecting data at IITH



Steps for inferential statistics

- Specify the question and identify the populations.
- How to select sample
- Select the sample and analyze the information
- Make the inference about the population
- Determine the reliability of inference

Descriptive statistics

Variable

Independent variable (x axis)

Dependent variable (y axis)

X

sin(x)

Qualitative

Example

Quantitative

Cases by Country Reported by MHLW as of April 19, 2020

	Infections	Fatalities
United States	726,645	30,664
Spain	191,726	20,043
Italy	175,925	23,227
Germany	143,160	4,459
Britain	114,717	15,464
France	111,821	19,323
China	82,735	4,632
Turkey	82,329	1,890
Iran	80,868	5,051
Russia	36,738	313
South Korea	10,661	234
Japan	10,361	161

Discrete

Continuous

Number of patient

weight

Students infected with COVID

height

distance

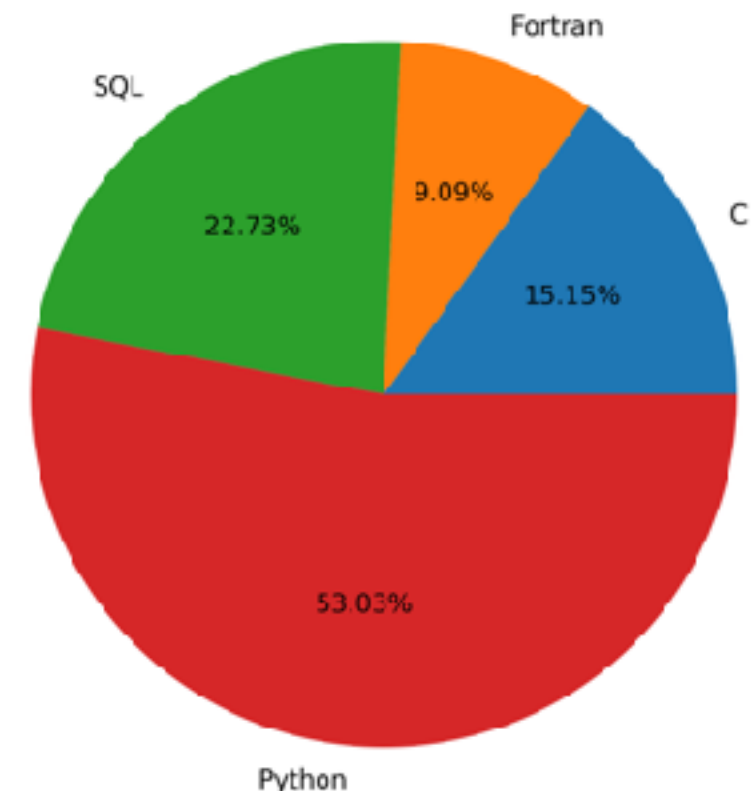
CGPA

Some plotting software

- GNUPLOT
- MATLAB
- MATHEMATICA
- XMGRC
- ORIGIN
- PYTHON (MATPLOTLIB PYPLOT)

Making “nice” plots

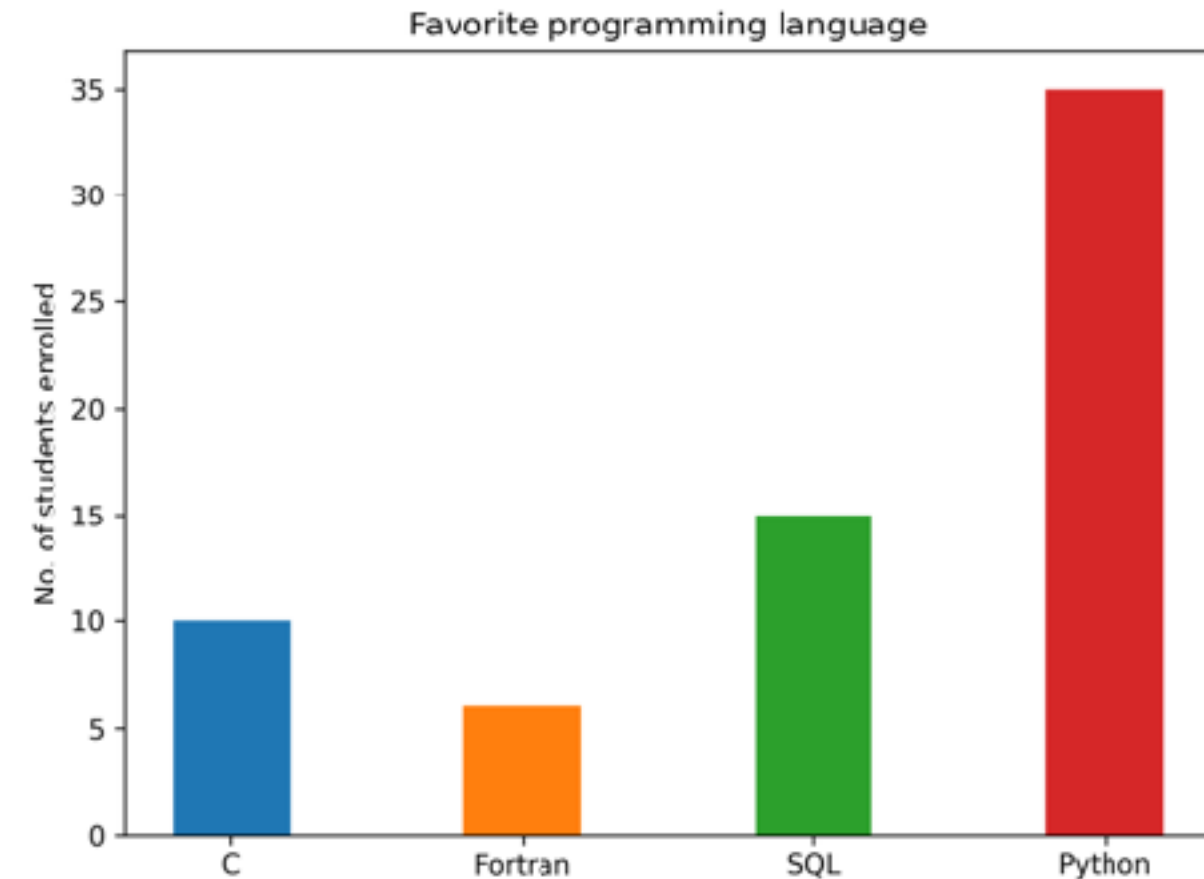
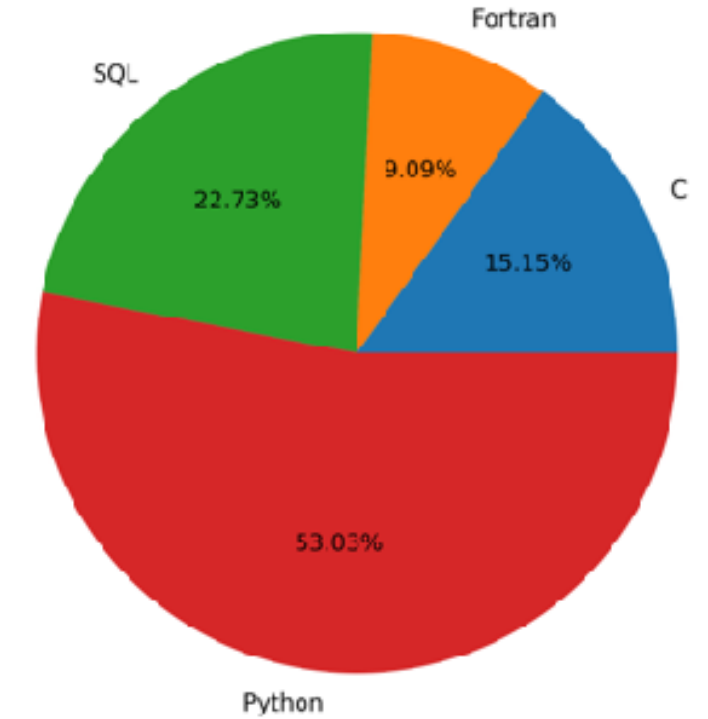
- Making an “informative, easy to interpret and attractive” plot is important, it makes your statement more clear and help during discussions, presentations etc.
- To solve the problem, the first step is to identify what the problem is
- What is wrong with the chart on the left





Converting data into figures

Favorite programming language	Number of students
C	10
Python	35
Fortran	6
SQL	15

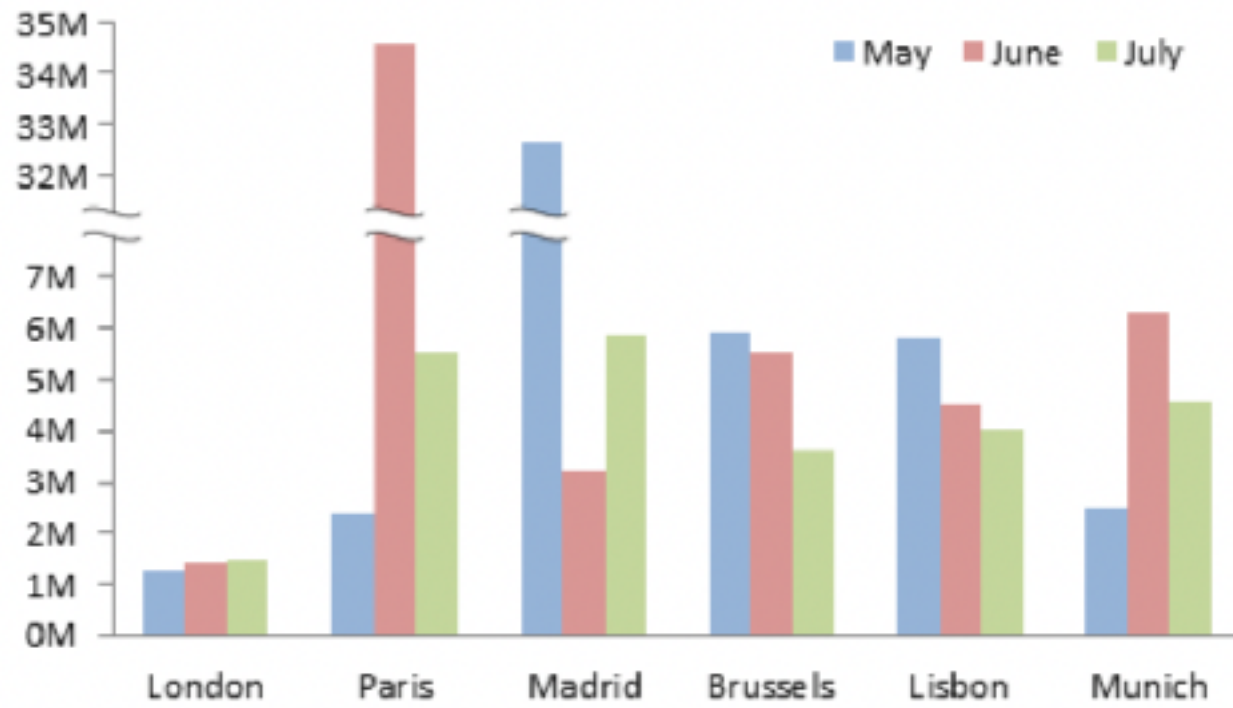


Problems with Pie chart or bar plot ?

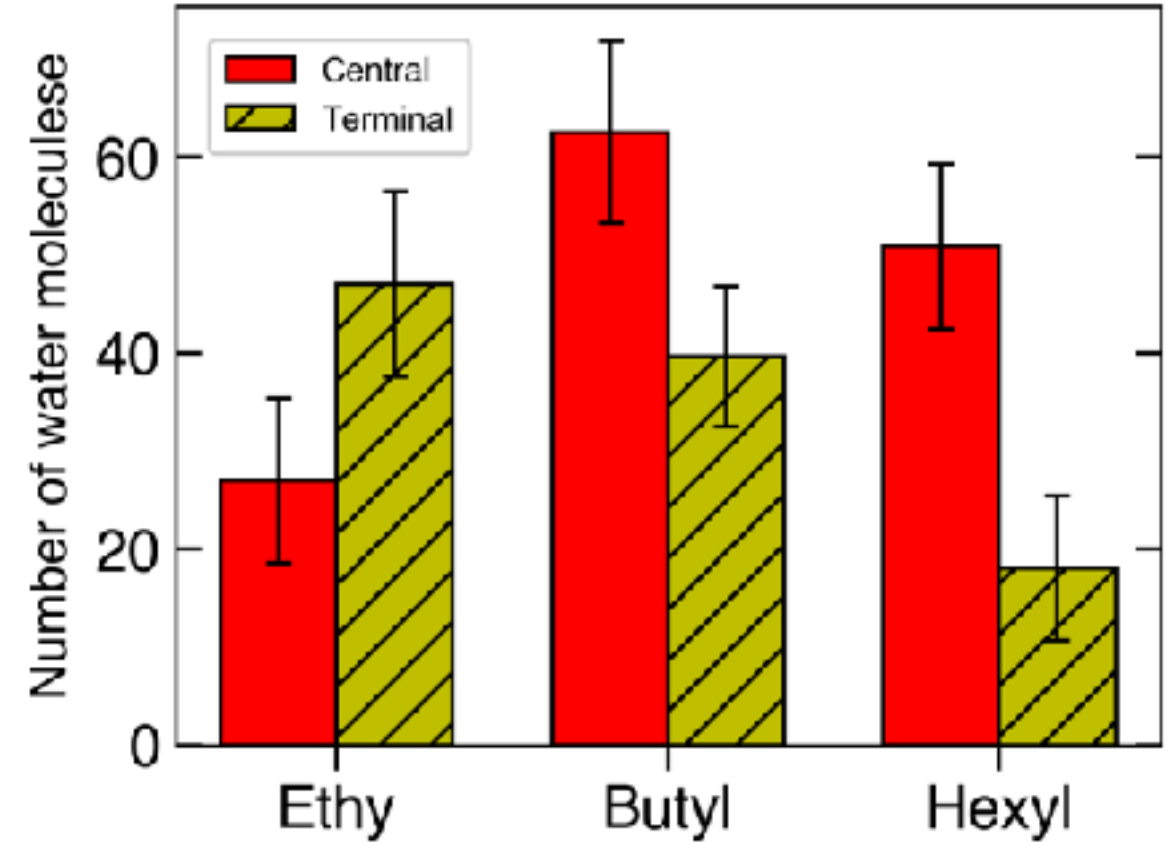


Some more about bar plot

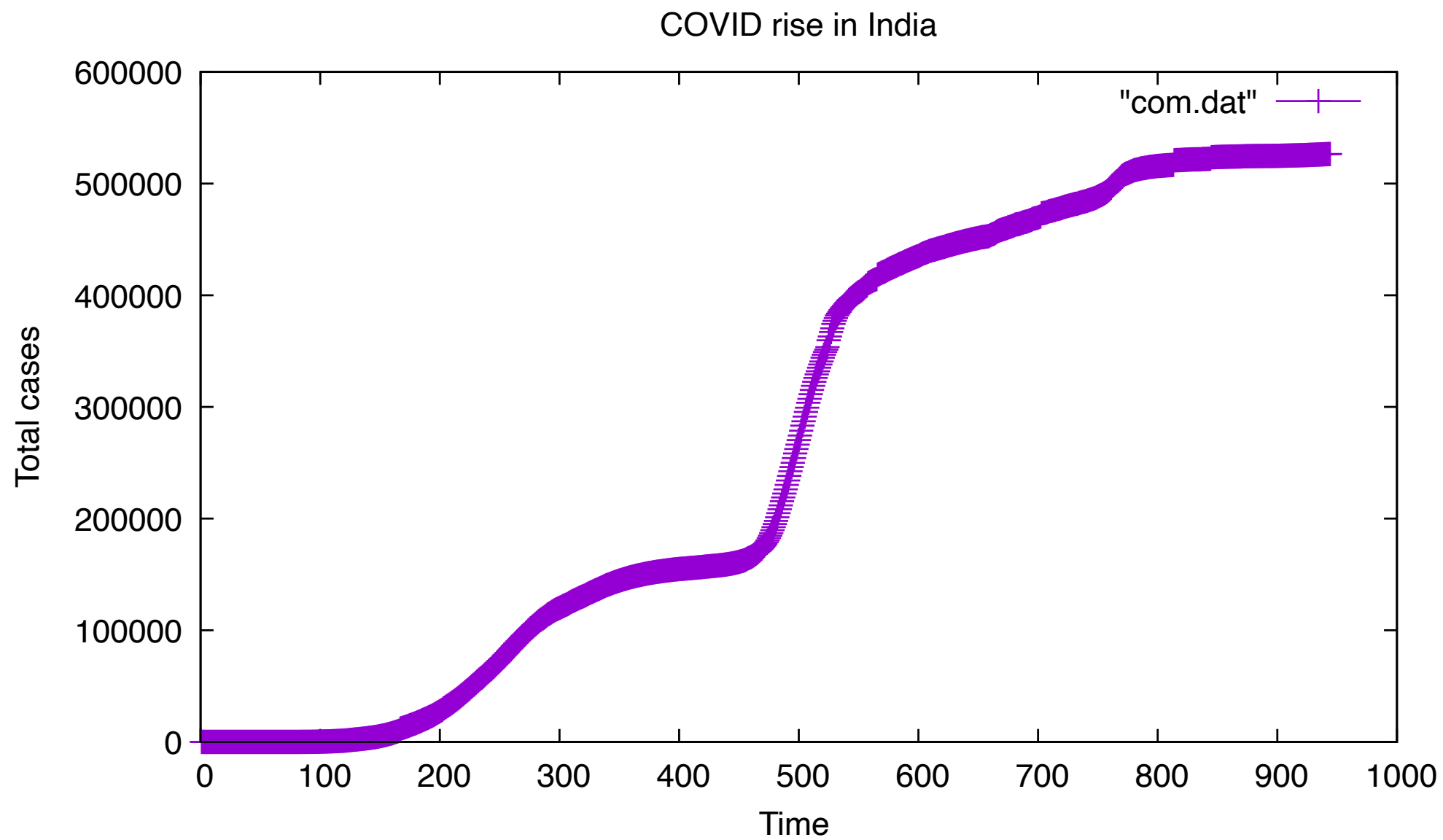
Broken axis



Error bars



Line plot



What is UNIX

- Introduced in 1969 in BELL labs (NOKIA), it is a multitasking operating system.
- The idea was to have a convenient platform for programmers developing software.
- **GNU** is an extensive collection of [free software](#); 383 packages until January 2022
- BASH (Bourne-Again SHell), GRUB, GCC, GFORTTRAN, GNPLOT etc
- Text editor “vi”
- sed, awk,

What is UNIX

- GUI vs command base interface
- There is always a computer code for GUI based interface
- Command based interface is suitable for programmers.
- Every action you do with mouse has a command version, so you see that programmers do not use mouse.

Name a command you use in GUI, we will try to see what is syntax

- pwd
- mkdir
- cd
- ls -ltr
- grep egrep
- find
- awk
- clear
- du -sh
- alias in bash
- cat
- >>
- history
- tab
- date

Basic UNIX commands

- cp -r
- rm -f
- mv
- ssh
- scp
- ipconfig
- rsync
- chmod
- any_command -h *or* -help
- tab *to complete*
- control r to back search

!!! Disclaimer : Do not try to memorize these commands

Save your memory to better things, its enough to know there exists a tool called sed, awk, cat, grep etc



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Next Class

2:30 PM Friday, 12 August 2022

Plotting and data visualization

continue

Python programming in Jupyter notebook !!