



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

Biostatistics BT2023

Lecture 2 and 3 : UNIX programming, GNU Plotting

Himanshu Joshi 08 August and 11 August 2023

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, GNUMPLOT etc
- Text editor “vi”
- sed, awk,

Text Editors

- **Notepad**
- **Emacs**
- **Vim**

Command mode

Insert mode

Replace mode

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

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
- expr

!!! 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

What is Truth

- **Belief:** A statement that is not scientifically provable in the same way as facts, laws, hypotheses or theories. Scientifically disproven beliefs can still be held to be true.
- **Hypothesis:** A tentative statement such as if **A** happens then **B** must happen that can be tested by direct experiment or observation. A proven hypothesis can be expressed as a law or a theory. A disproven hypothesis can sometimes be re-tested and found correct as measurements improve.
- **Conjectures:** An idea, hypothesis is a conjecture which can be tested.
- **Fact:** A basic statement established by experiment or observation. All facts are true under specific conditions. Some facts may be false when re-tested with better instruments.
- **Law:** A logical relationship between two or more things that is based on a variety of facts and proven hypothesis. It is often a mathematical statement of how two or more quantities relate to each other.
- **Theory:** An explanation for why certain laws and facts exist that can be tested to determine its accuracy.

Python

Python is an object oriented high level programming language
Conceived in 1980 in Netherlands.

Objects

Data

Procedure



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

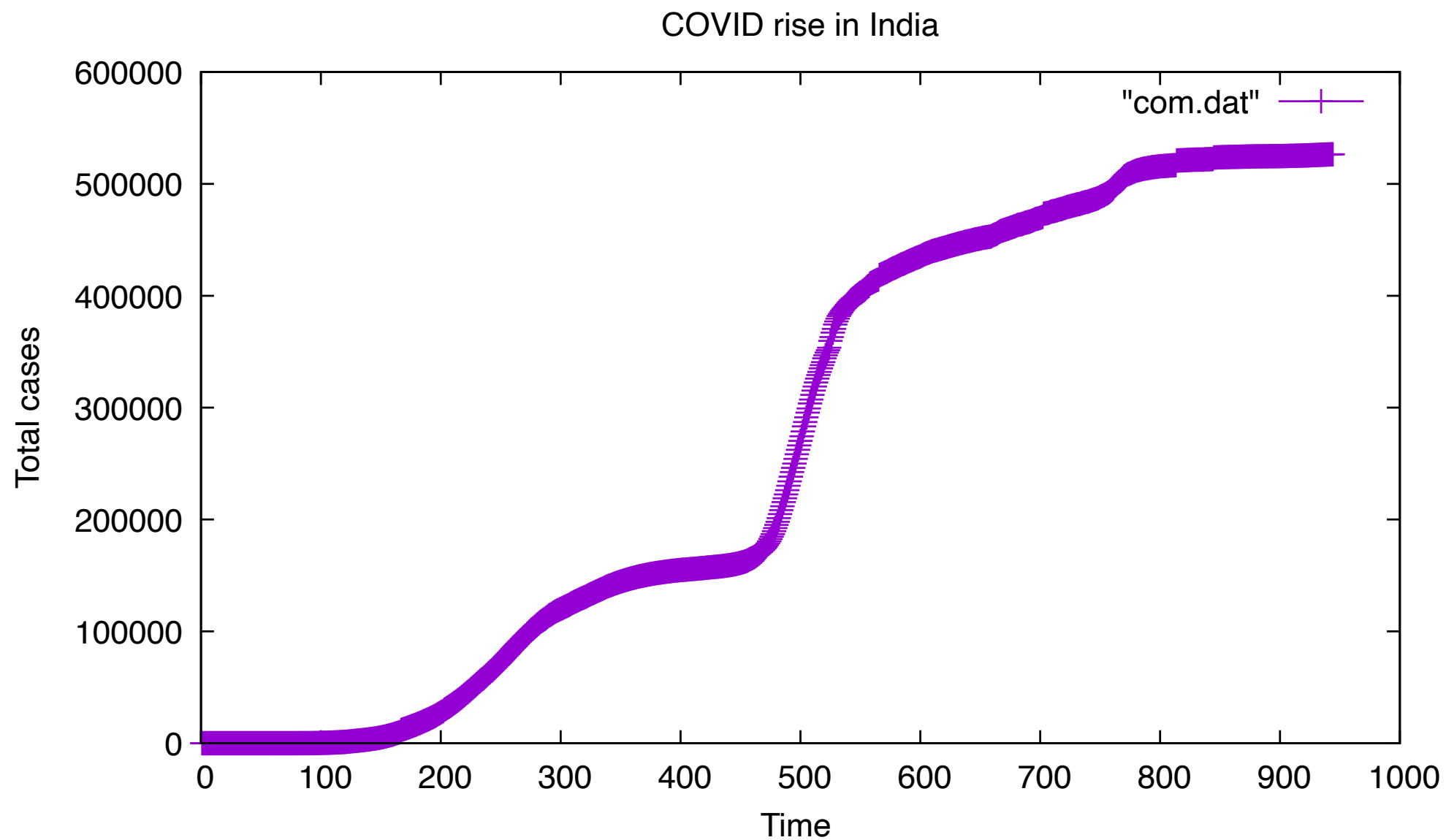
distance

CGPA

Some plotting software

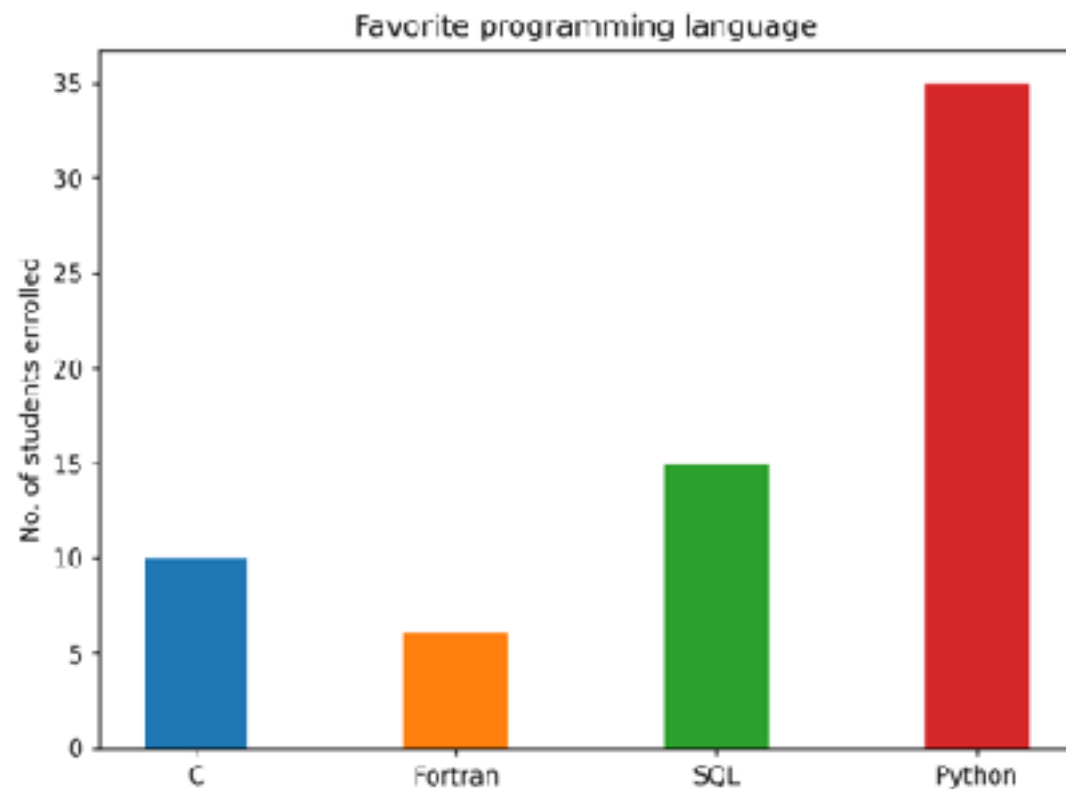
- GNUPLOT
- MATLAB
- MATHEMATICA
- XMGRC
- ORIGIN
- PYTHON (MATPLOTLIB PYPLOT)

Line plot



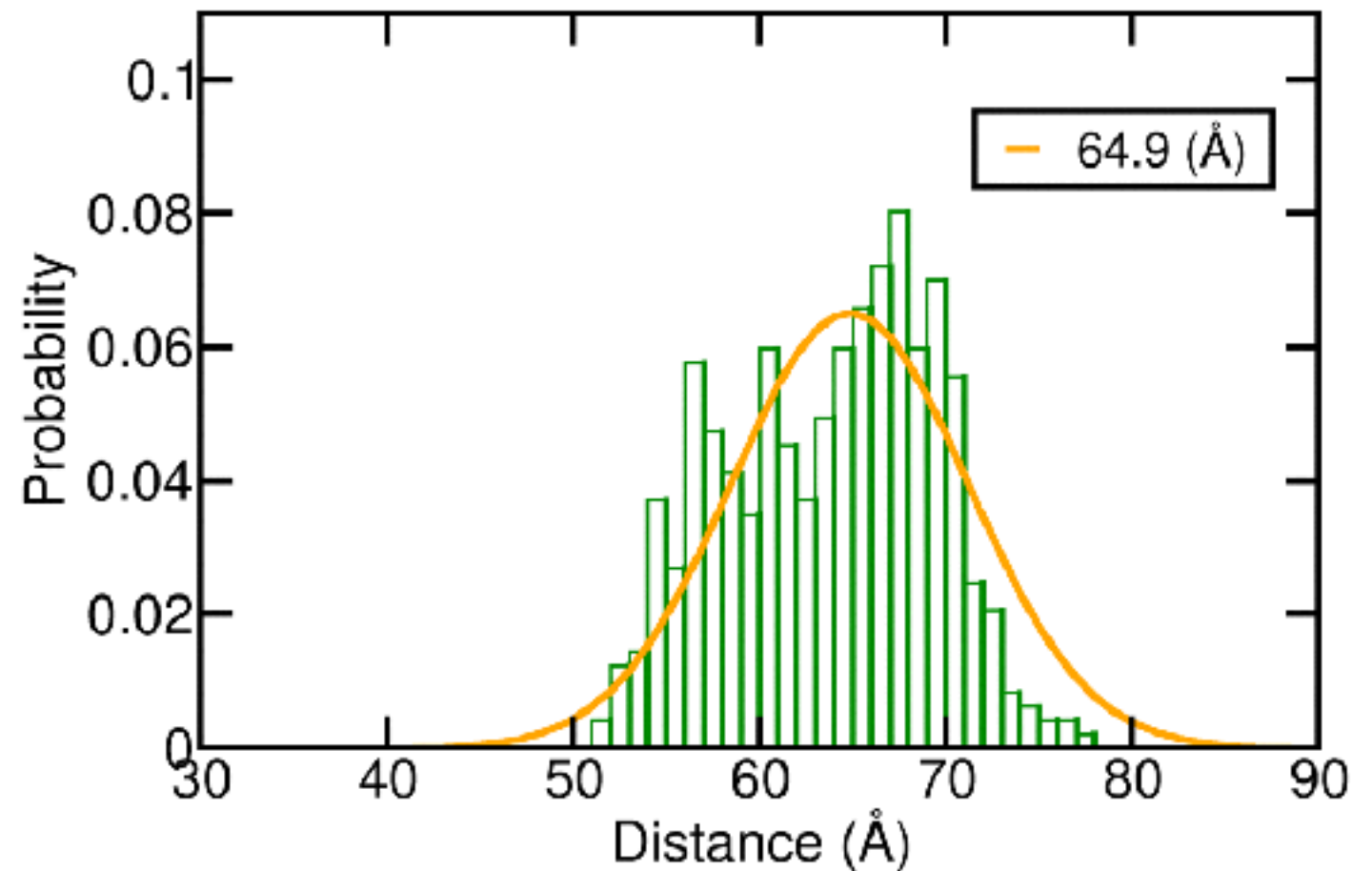
Histogram and bar plot

Bar plot



diagrammatic comparison of
discrete variables

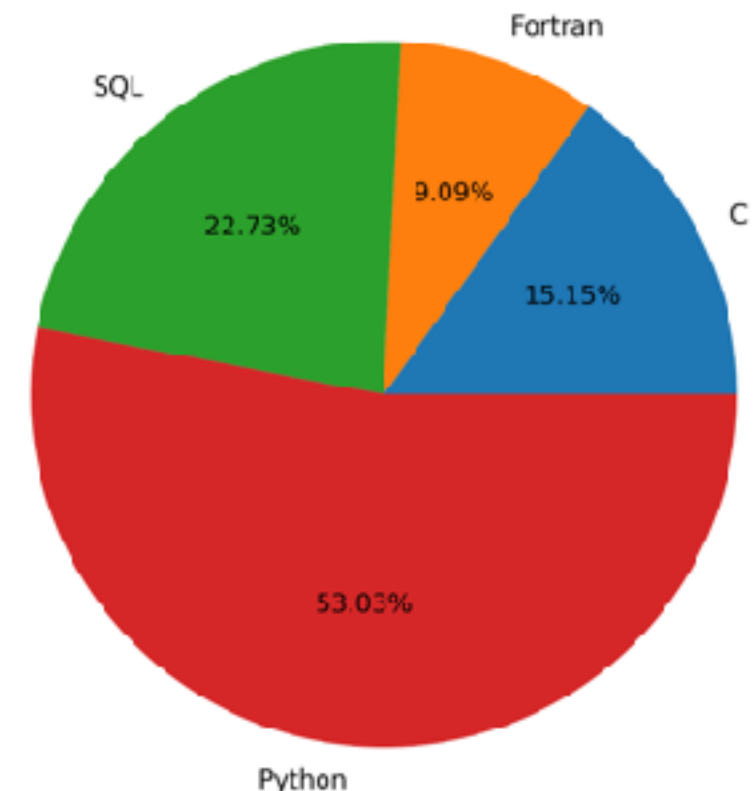
Histogram



frequency distribution of
continuous variables.

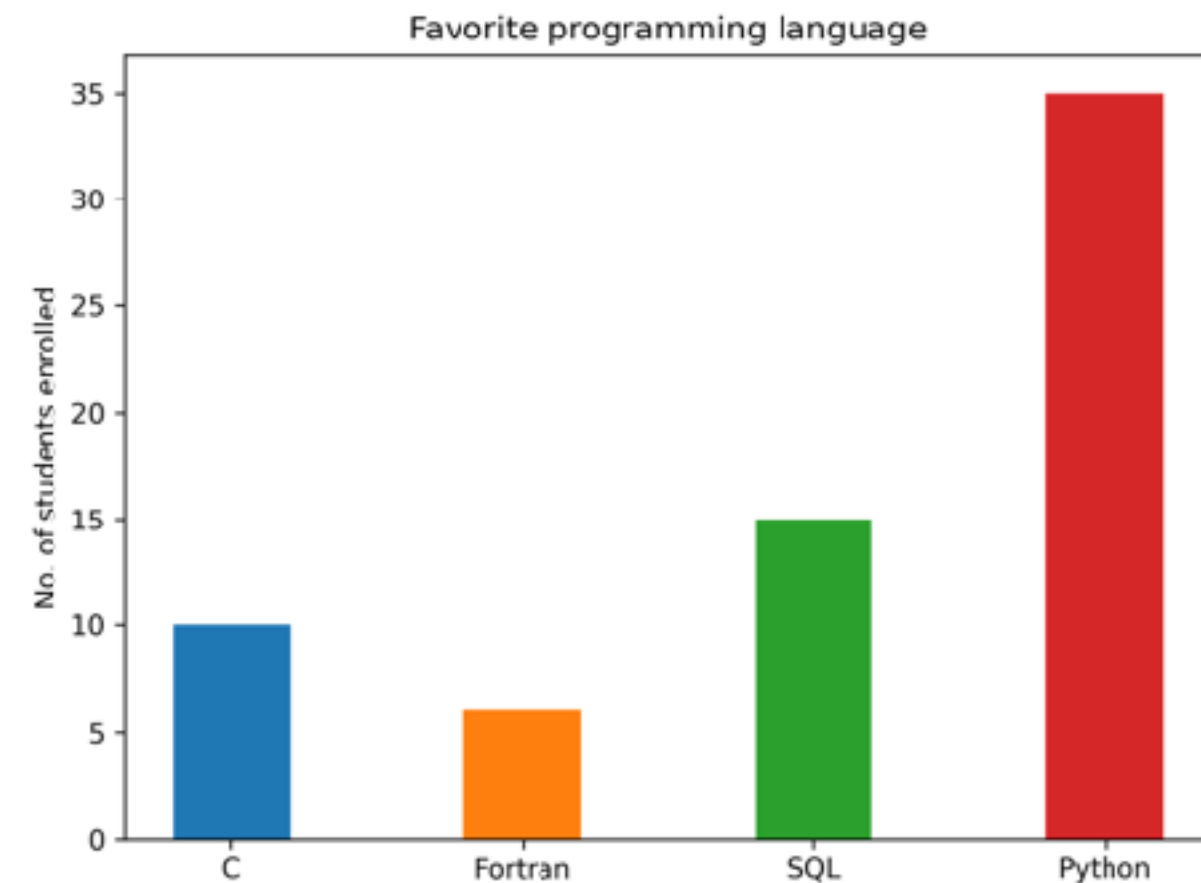
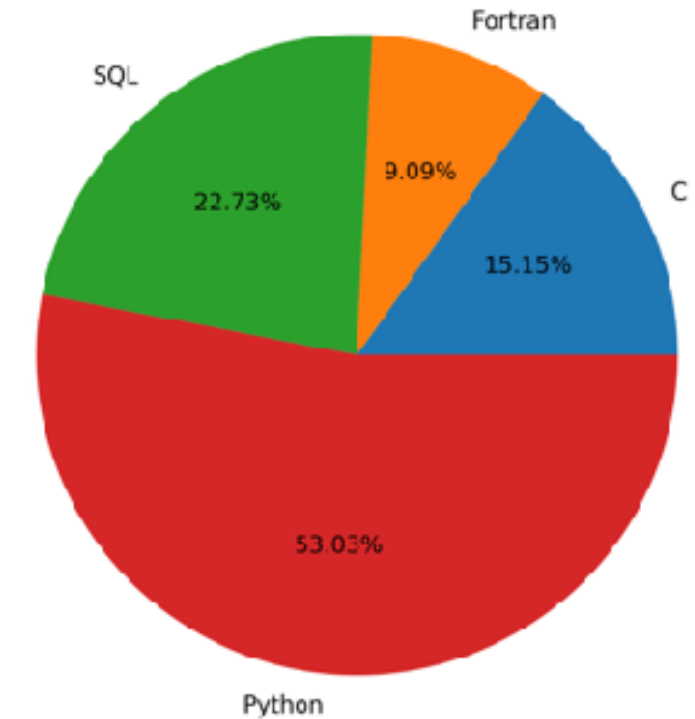
Making “nice” plots

- 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

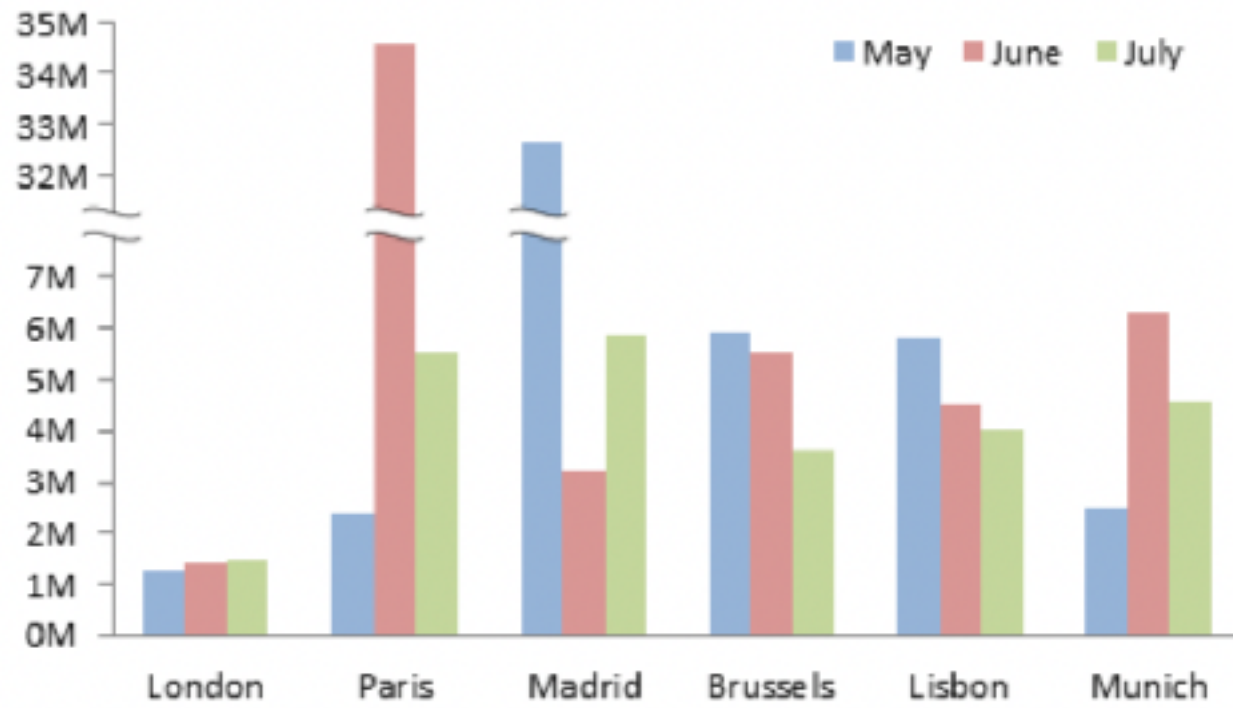


Problems with Pie chart or bar plot ?

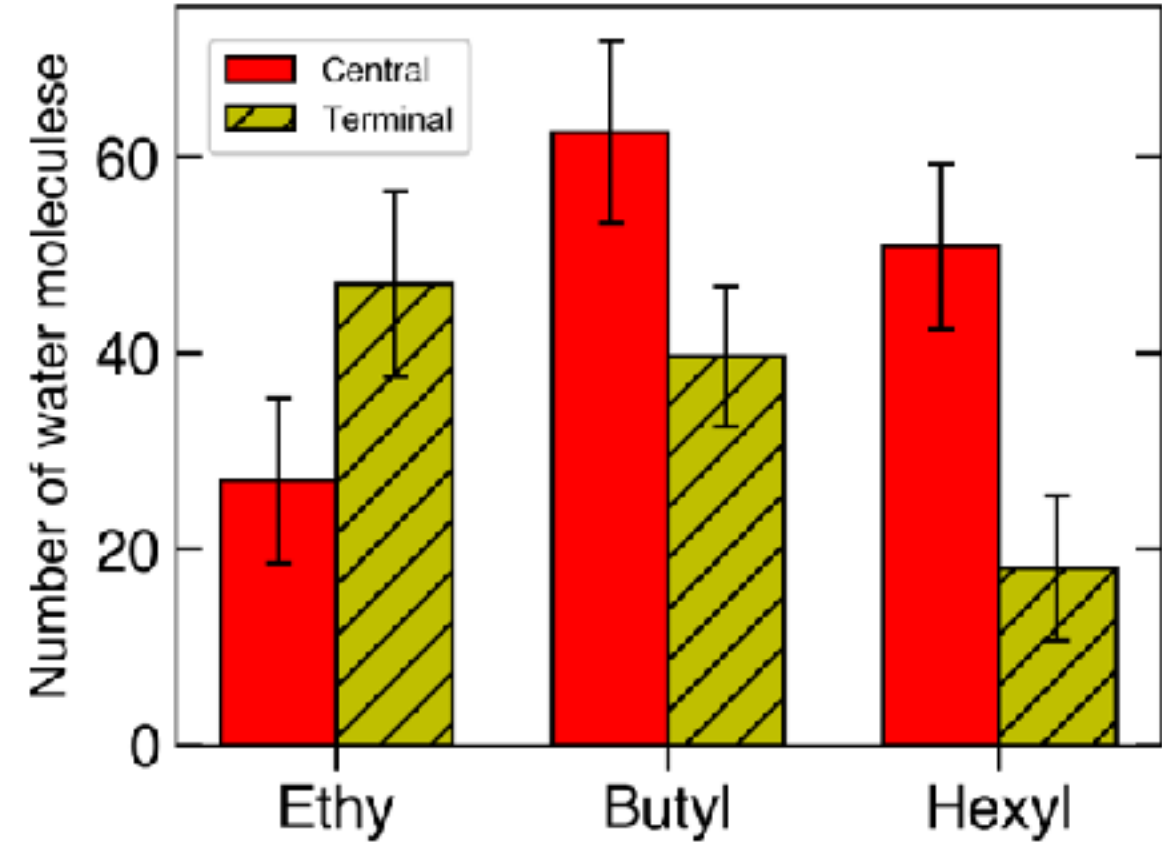


Some more about bar plot

Broken axis



Error bars



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



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

Next Class

2:30 PM Friday, 11 August 2022

Plotting and data visualization

continue

Python programming