Introduction to Mathematical Foundations for Data Science-I (MAT 2157)

- *Topics covered:*

✓ **Data types**
✓ **Frequency tables**
✓ **Graphical representation**

Dr. HARINAKSHI KARKERA

MAT 2157: MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE-I [3 0 0 3]

Statistics: Measures of central tendency – mean, median, mode, measures of dispersion – mean deviation, standard deviation, quartile deviation, skewness and kurtosis. Correlation coefficient, regression, least squares principles of curve fitting. 8

Probability: Introduction, finite sample spaces, conditional probability and independence, Baye's theorem, one dimensional random variable, mean, variance. Two and higher dimensional random variables: mean, variance, correlation coefficient.

Distributions: Binomial, Poisson, uniform, normal, gamma, Chi-square and exponential distributions, simple problems. Moment generating function, Functions of one dimensional and two dimensional random variables, Sampling theory, Central limit theorem and applications.

Partial Differential Equations: Numerical solutions of partial differential equations by finite difference methods, five-point formula, Laplace Poisson Equations, Heat equation, Crank Nicolson's method, Wave equation. Application of Finite Difference technique.

References:

1. Meyer P.L. "*Introduction to probability and statistical applications*", 2nd edn., American Publishing Co
2. Hogg and craig , *Introduction to mathematical statistics*, 6th Edn,2012, Pearson education, New Delhi.
3. Ross Sheldon M, "*Introduction to Probability and Statistics for Engineers and Scientists*", Elseveir, 2010.
4. William J. Stewart, "*Probability, Markov Chains, Queues and Simulation*".
5. S. Narayanan, T. K. Manicavachagom Pillay, G. Ramanaiah, *Advanced mathematics for engineering students*, S. Viswanathan Pvt.. Ltd., 1985.

Gupta, S.C and Kapoor, V.K. Fundamentals of Mathematical Statistics, Sultan Chand & Sons,11th Edition, 2016

Course Outcomes (COs)

	At the end of this course, the student should be able to:	No. of Contact Hours	Marks
CO1:	Understand the basic concepts of descriptive statistics in any data set	08	10
CO2:	Understand the various probability laws in any real data	10	10
CO3:	Identify the dependency structure among variables using correlation coefficient and build models for forecasting	05	10
CO4:	Test for consistency and validate the data using the concepts in sampling theory	06	10
CO5:	Solve partial differential equations using numerical methods	05	10
	Total	34	50

Lesson Plan

L. No./ T. No.	Topics	Course Outcome Addressed
L0	Introduction to the subject stating its importance	CO1
L1	Data types, frequency tables, graphical representation	CO1
L2	Measures of central tendency-numerical and graphical	CO1
L3	Measures of central tendency-numerical and graphical	CO1
L4	Measures of variation-relative measures	CO1
L5	Measures of variation-coefficient of variation	CO1
T1	Tutorials	CO1
L6	Raw moments, central moments	CO1
L7	Measures of skewness, kurtosis	CO1
L8	Analyse data using the concepts learnt lectures 1 to 7	CO1

Dr.

Variable ✓

Measurements ✓

Data : Data are recorded measurement

Humans,
Machines,
H-M

Dr. HARINAKSHI KARKERA

Why data is important?

* To make better decisions
* To have a better understanding of a certain issue
* To find hidden patterns
* To predict the future
* To evaluate the performance
* To improve services

Dr. HARINAKSHI KARKERA

News Flash: Everything you need to know about Zee's merger with Sony

Data science the most coveted domain by both students and professional developers: Survey

By Brinda Sarkar, ET Bureau • Last Updated: Jul 26, 2021, 03:53 PM IST

Synopsis
Other domain skills that student developers would like to explore in future include cybersecurity (16.1%), game development (13.8%), Internet of Things (13.4%), blockchain (12.6%), robotics (9.5%), and business intelligence (9.2%), according to the survey by HackerEarth, a global tech hiring platform for enterprises.

Lenovo device-to-d
center solutions.
Learn more

Dr.

The survey finds that YouTube is a popular way to upskill among both students and professionals.

Data science has emerged as the domain that is most coveted by both students and professional developers, with approximately 25% respondents in a survey pointing to it as a domain skill they would like to learn in future.

Other domain skills that student developers would like to explore in future include cybersecurity (16.1%), game development (13.8%), Internet of Things (13.4%), blockchain (12.6%), robotics (9.5%), and business intelligence (9.2%), according to the survey by **HackerEarth**, a global tech hiring platform for enterprises.

The 'Code in Progress: HackerEarth 2021 Developer Survey' has 25,400-odd responses from developers across 171 countries with over 20% of the respondents being women. The survey aims to shed light on what the programming world looks like through a developer's lens with insights into

The Indian EXPRESS

Wednesday, September 22, 2021

Diversity and access can lead many more Indians to jobs in data science

Alongside knowledge of standard techniques for analysis, your success as a data scientist depends on your ability to interpret data, innovate and bring a creative approach to problem-solving.

New Delhi |
Updated: March 7, 2021 9:57:40 am

— *Written by Emily Glassberg Sands*

By 2025, the top job role in the world — with the most growth in demand — will be that of a data analyst and scientist, according to the World Economic Forum's Future of Work Report 2020. As the demand for data scientists outstrips supply worldwide, India is grappling with its own serious shortage, according to industry reports. The country also trails on data science skills — ranking at 51 out of 60 countries in Coursera's Global Skills Index Report 2020.

At the same time, India has the opportunity to significantly expand its pipeline of talent, if many more students across disciplines prepare for jobs in data science. Not every data scientist has a STEM background or needs one. Data scientists can come from different fields — I'm a labour economist by training and first applied data science for my PhD thesis. It is not uncommon for companies globally to fill out their data science teams with talent from a broader choice of fields — including social sciences — alongside traditional hires like computer scientists.

Types of data

1) Qualitative] ✓

2) Quantitive] — numerical
   - Discrete ← No. of children
   - Continuous ← temperature, height, voltage

---

1) Nominal } lowest level of measurement
2) Ordinal
3) Interval
4) Ratio } Highest level of measurements

Dr. HARINAKSHI KARKERA

1) Nominal : Gender < 0 M, 1 F }

Primary data

Secondary data

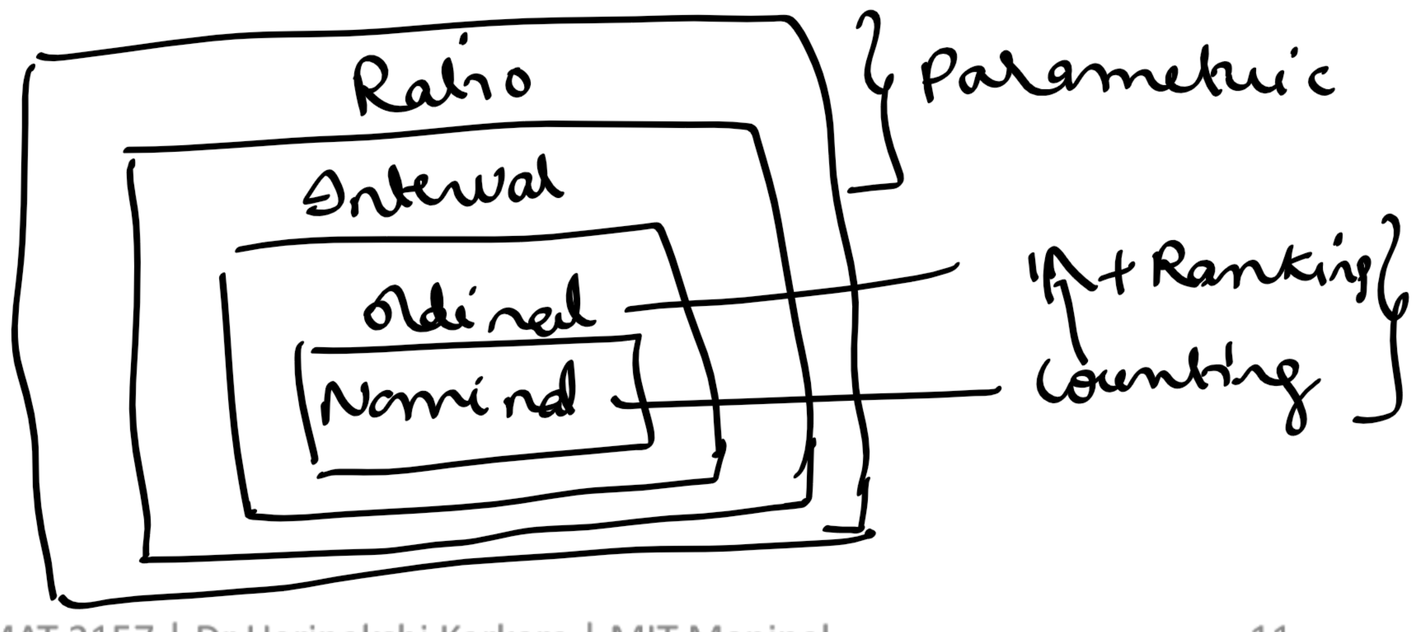
2) Ordinal : Faculty rank

3) Interval : Ex: temperature '0'F freezing pt

year $2021 - 2000$

$= \underline{\underline{21}}$

4) Ratio :



Dr. HARINAKSHI KARKERA

Frequency distribution

Demo_MarksDistribution - Excel

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do...

K12

	A	B	C	D	E	F	G
1	Sl. No.	Name	MCQ (out of 12)	Descriptive-2 Marks each (out of 16)	Descriptive-3 Marks each (out of 9)	Total (Out of 37)	
2	1	KAZI	11	15.5	9	35.5	
3	2	ARPIT	12	15	6	33	
4	3	PRATHAMESH	12	14	7	33	
5	4	SAAKETH	11	16	7	34	
6	5	SUJITHA	12	14.5	6	32.5	
7	6	JAYSHREE	12	11.5	6	29.5	
8	7	DEVIKA	12	10.5	7.5	30	
9	8	SRIVASTAV	8	13	6.5	27.5	
10	9	REET	8	8.5	5	21.5	
11	10	RUDRANI	12	13.5	7	32.5	
12	11	PRASHANTH	12	12	9	33	
13	12	SHAGUN	11	14	7.5	32.5	
14	13	SOHAM	11	9.5	5	25.5	
15	14	HARISH	9	5.5	8.5	23	
16	15	TEJASWINI	12	15	7	34	
17	16	MADHU	9	15.5	6	30.5	
18	17	VINOD	12	13	4	29	
19	18	SUHAS	10	13.5	9	32.5	
20	19	YASH	12	15	6	33	
21	20	JOHN	12	13.5	6	31.5	
22							

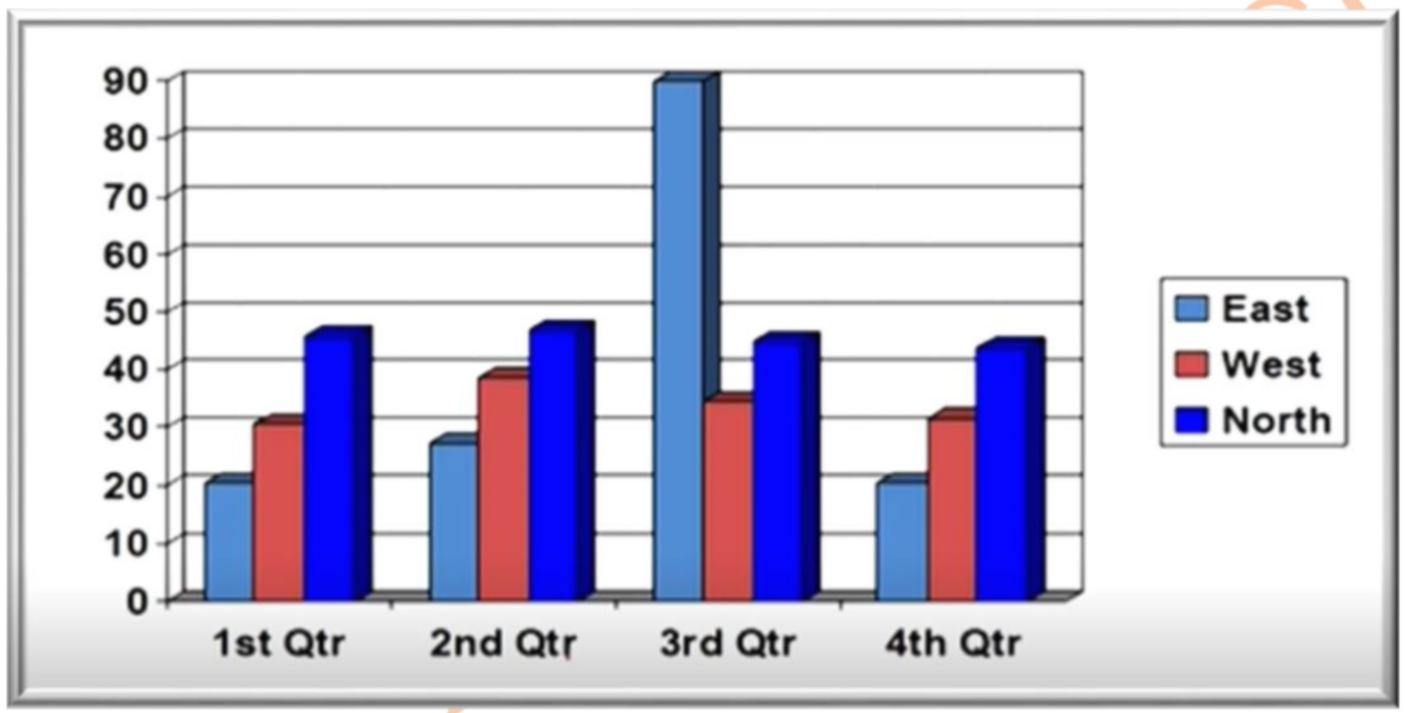
'x' Marks	(Tally Marks) No. of students	'f' frequency
21.5	1	1
22	0	0
22.5	0	0
23	1	1
23.5	⋮	⋮
32.5	1111	4
33	1111	4
35	0	0
35.5	1	1
		20

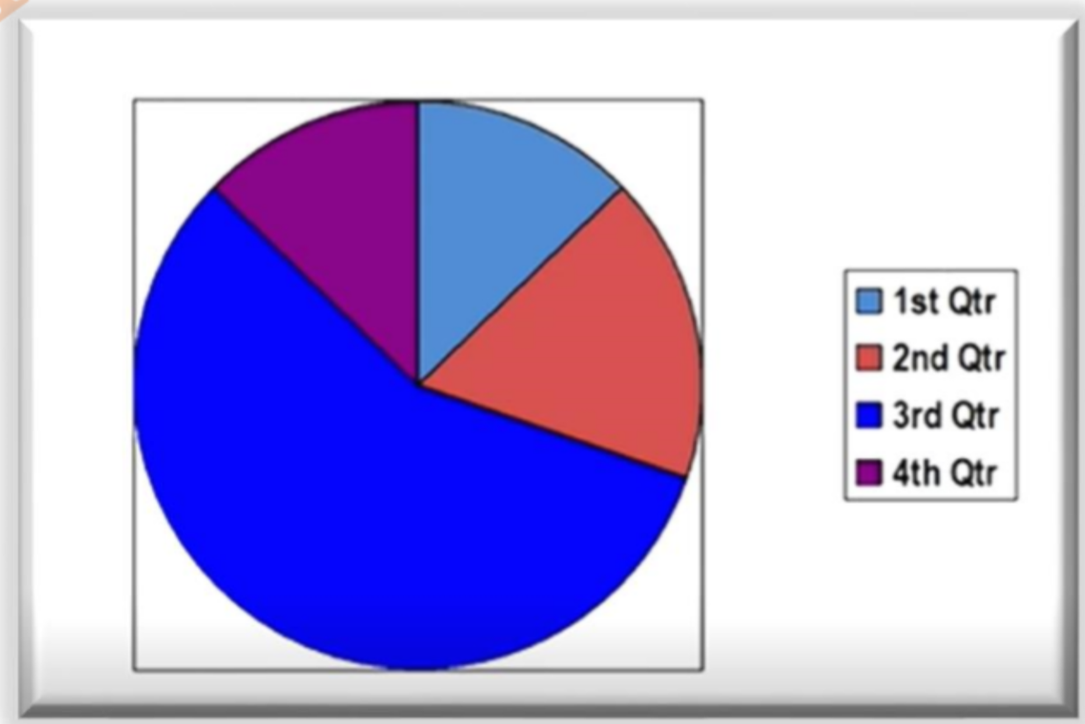
frequency

Dr. HARINAKSHI KARKERA

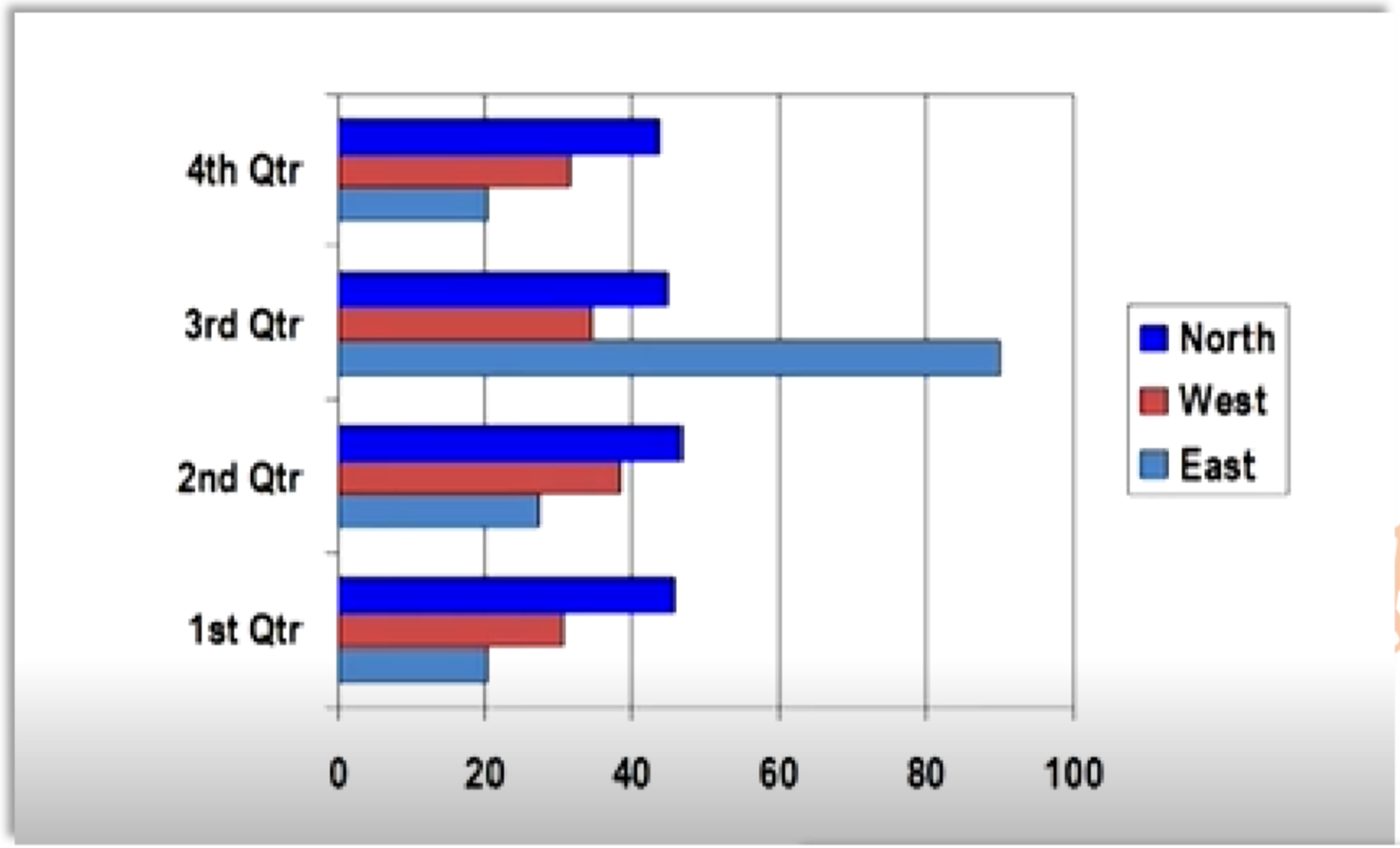
Graphical representation

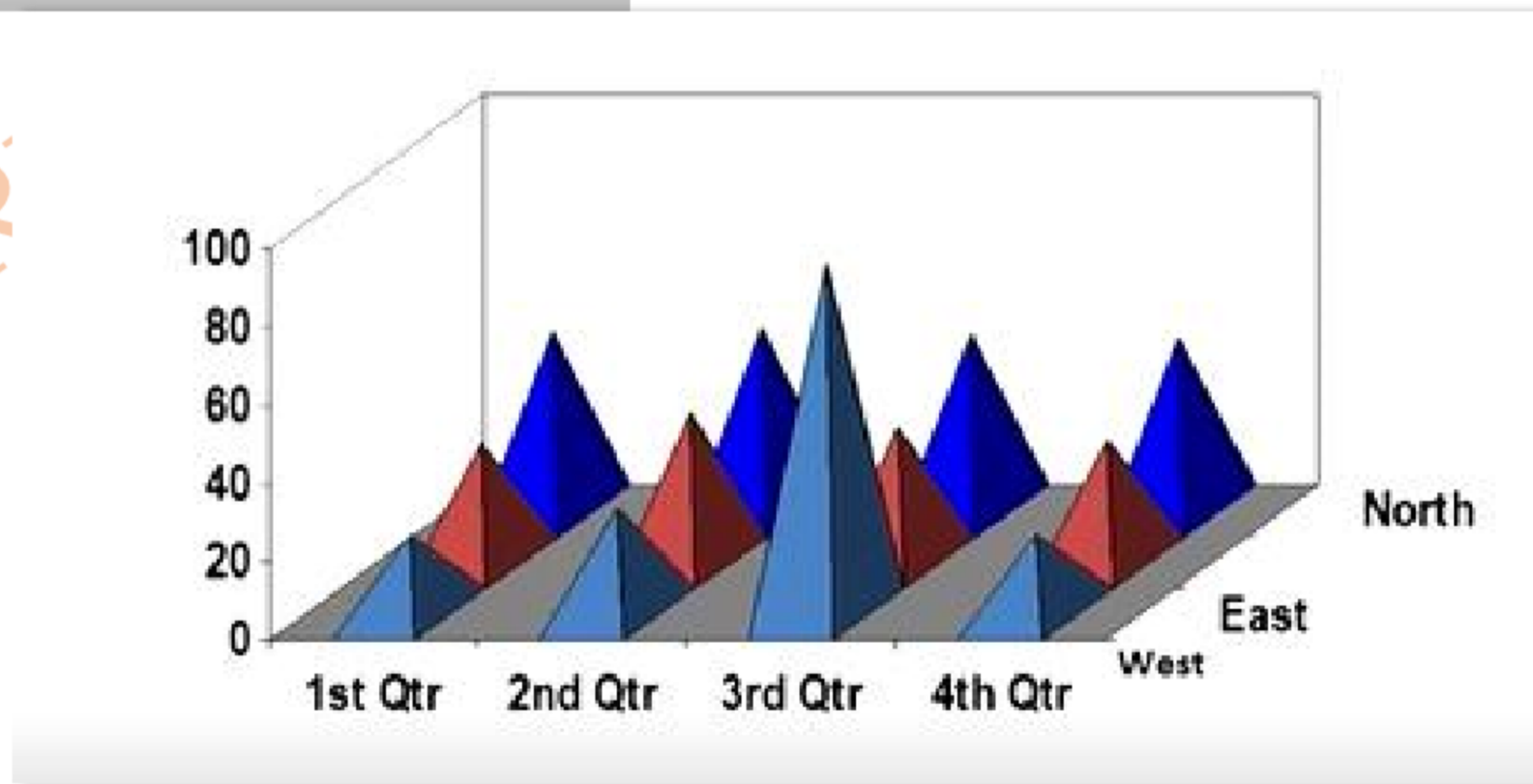
	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr
East	20.4	27.4	90	20.4
West	30.6	38.6	34.6	31.6
North	45.9	46.9	45	43.9





ARKERA





Dr. HARKARKERA

Graphical representation of a frequency distribution

?

Dr. HARINAKSHI KARKERA