



Start-TechAcademy

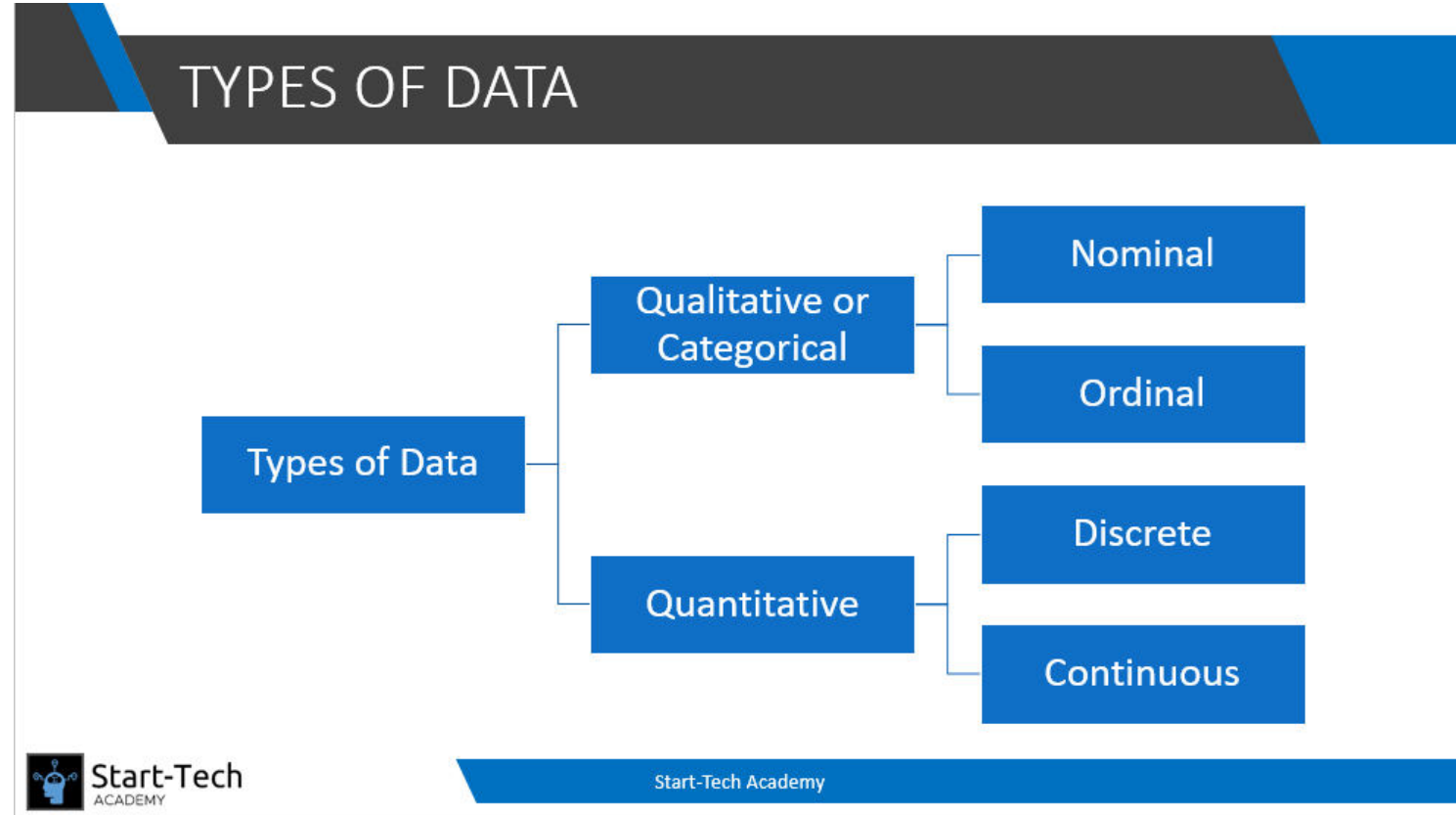


Machine Learning: Building a Linear Regression model

Abhishek & Pukhraj

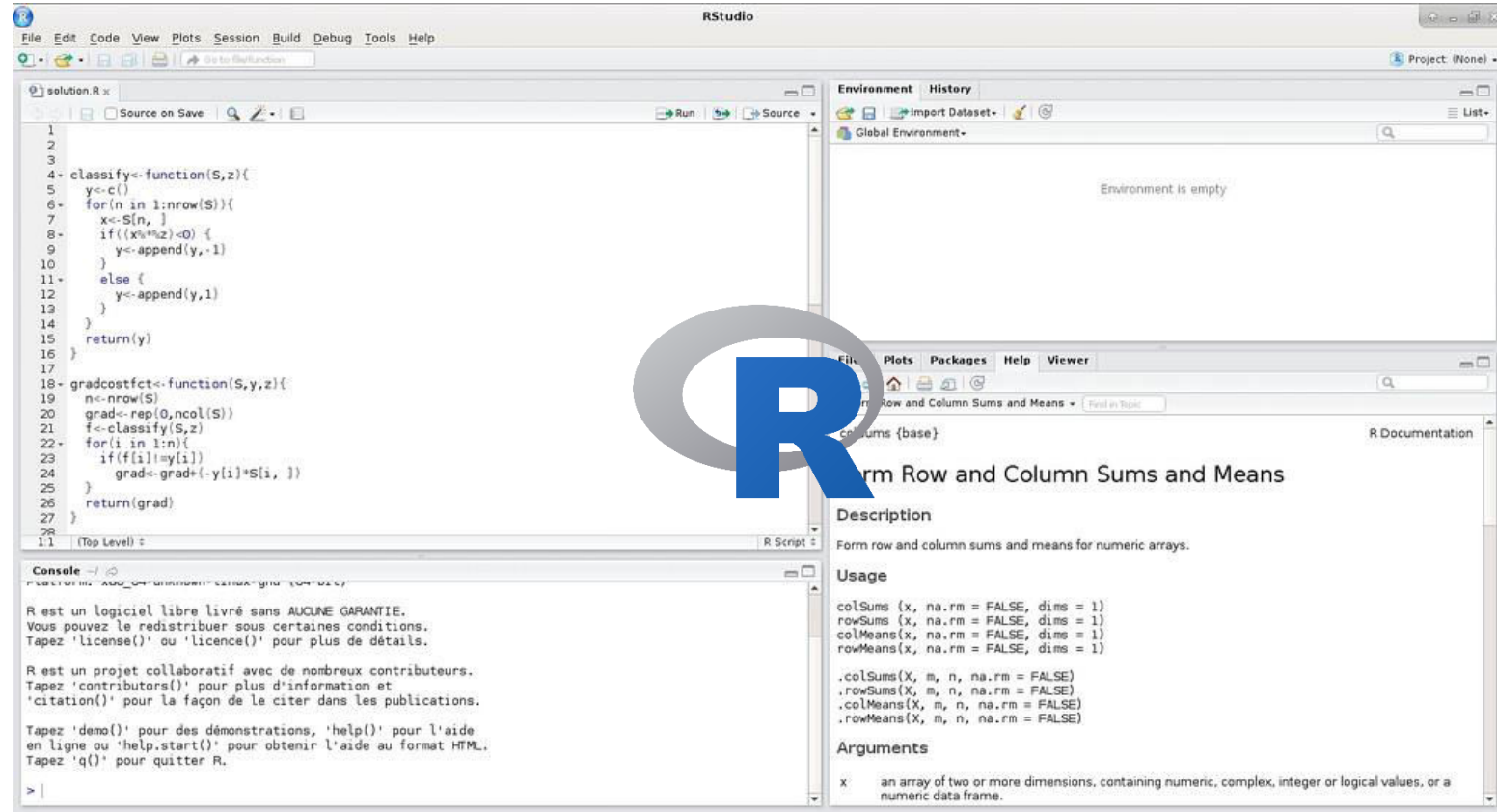
Basics of Statistics

- ✓ Types of Data
- ✓ Types of Statistics
- ✓ Describing Data graphically
- ✓ Measures of Center and Dispersion



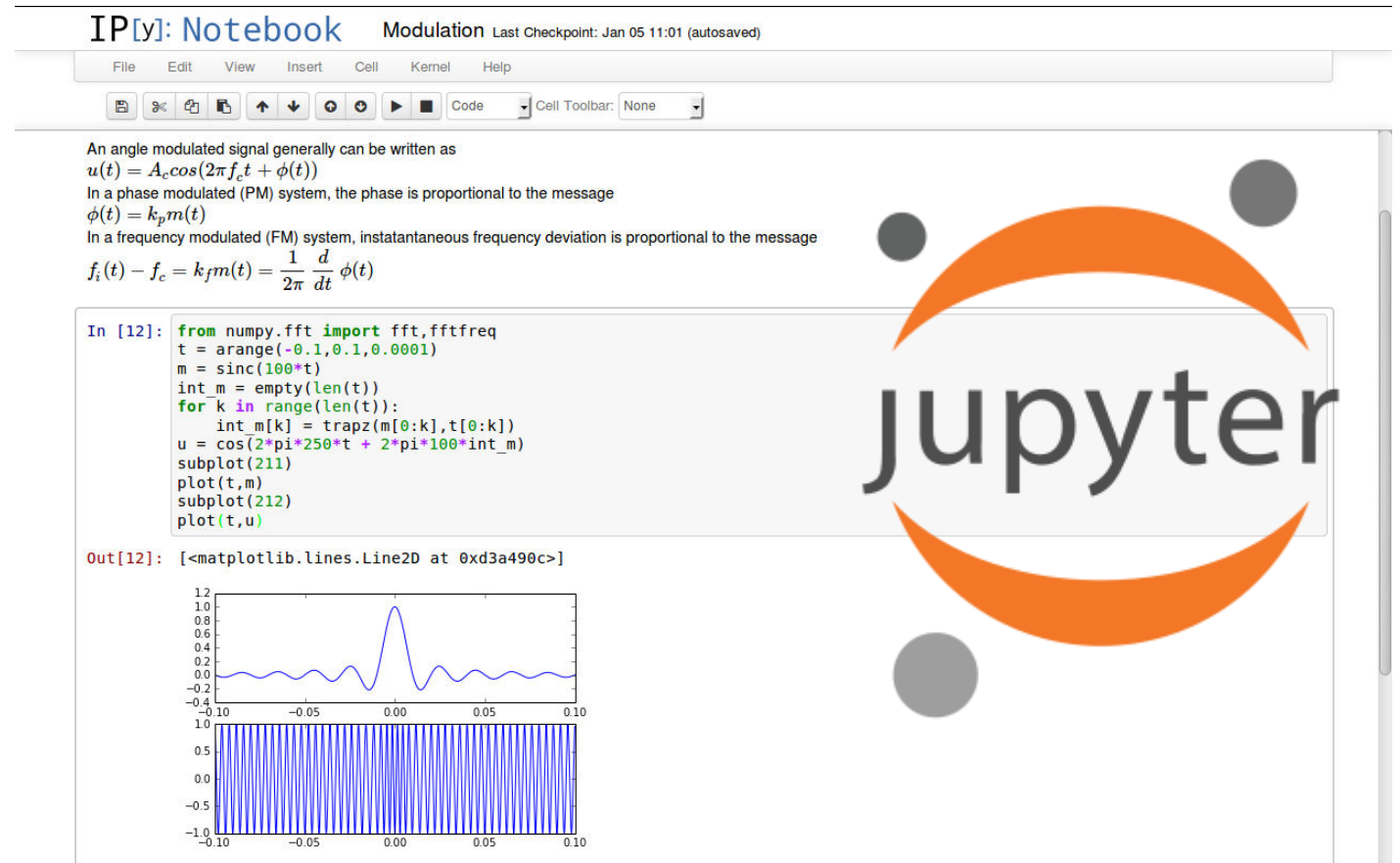
Getting Started with R / R Studio

- ✓ Installation
- ✓ Basic Operations
- ✓ Packages
- ✓ Importing Data
- ✓ Bar plots
- ✓ Histograms



Getting Started with Python

- ✓ Installation
- ✓ Jupyter Notebook
- ✓ Python Basics
- ✓ Importing Data
- ✓ Numpy / Pandas
- ✓ Seaborn



Machine Learning

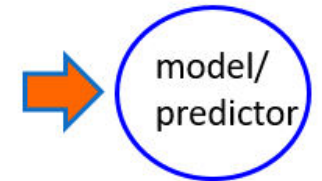
- ✓ Introduction to Machine Learning
- ✓ Building a Machine Learning Model

Supervised Learning: Example

Supervised
Learning
Example



Category	Weight
Apple	100 gm
Apple	80 gm
Banana	40 gm
Banana	60 gm



Data Preprocessing – Part 1



- ✓ Business Knowledge
- ✓ Data Exploration
- ✓ Univariate Analysis
- ✓ Bivariate Analysis

	price	crime_rate	resid_area	air_qual	room_num	age	teachers	poor_prop	n_hos_beds
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
mean	22.528854	0.813418	41.136779	0.554695	6.284634	68.574901	21.544466	12.653063	7.899767
std	9.182176	1.022731	6.860353	0.115878	0.702617	28.148861	2.164946	7.141062	1.464939
min	5.000000	0.006300	30.460000	0.385000	3.561000	2.900000	18.000000	1.730000	5.268000
25%	17.025000	0.078853	35.190000	0.449000	5.885500	45.025000	19.800000	6.950000	6.659000
50%	21.200000	0.228336	39.690000	0.538000	6.208500	77.500000	20.950000	11.360000	7.963000
75%	25.000000	1.542674	48.100000	0.624000	6.623500	94.075000	22.600000	16.955000	9.076000
max	50.000000	4.499545	57.740000	0.871000	8.780000	100.000000	27.400000	37.970000	10.876000

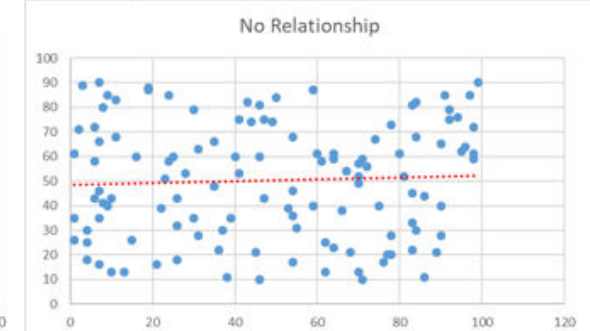
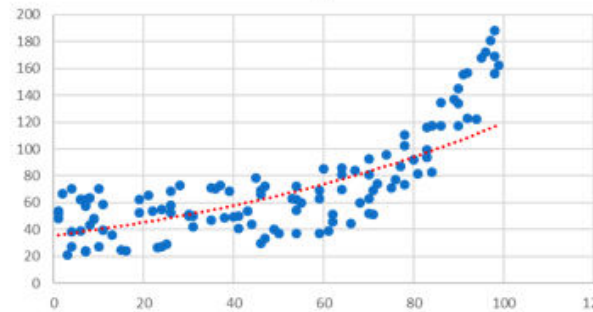
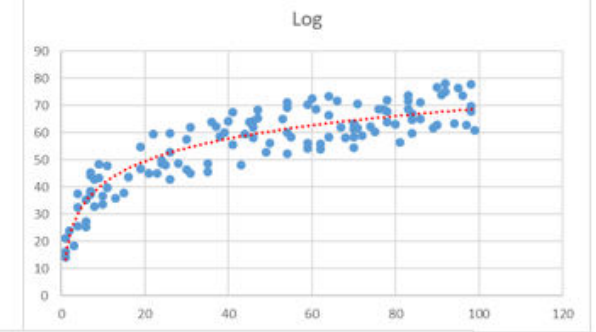
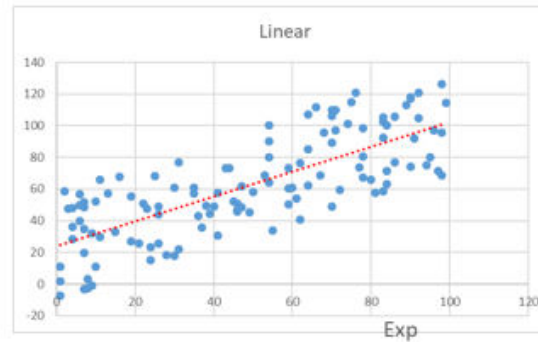


Data Preprocessing – Part 2



- ✓ Outlier Treatment
- ✓ Missing Value Imputation
- ✓ Variable Transformation
- ✓ Correlation

Scatter plots



Linear Regression – Part 1



- ✓ Ordinary Least Squares
- ✓ Simple Linear Regression
- ✓ Multiple Linear Regression

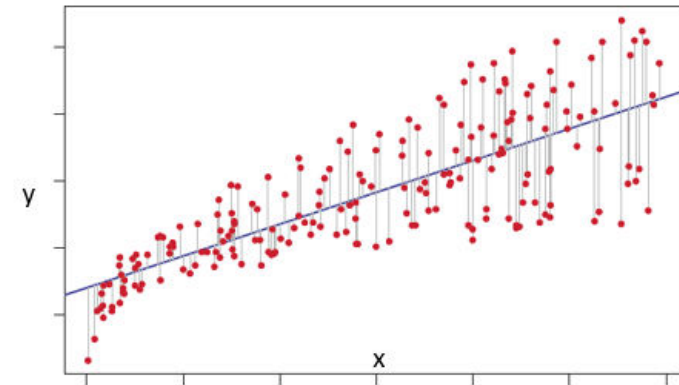
Simple Linear Regression

Residual –

The difference between residual the i th observed response value and the i th response value that is predicted by our linear model is known as residual

$$e_i = y_i - \hat{y}_i$$

Residual



Linear Regression – Part 2



- ✓ Model Accuracy
- ✓ F Statistics
- ✓ Categorical Variable Results
- ✓ Other Variations
- ✓ Result Interpretation

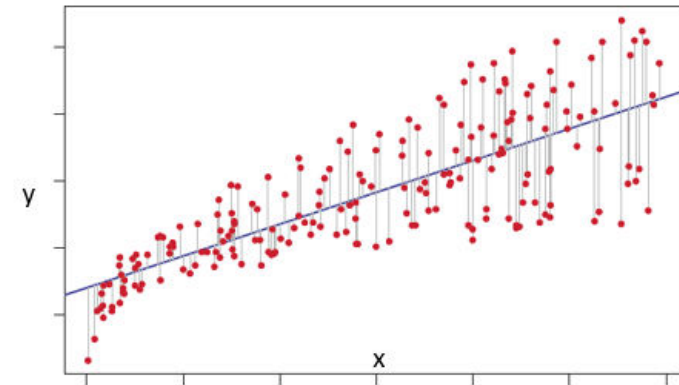
Simple Linear Regression

Residual –

The difference between residual the i th observed response value and the i th response value that is predicted by our linear model is known as residual

$$e_i = y_i - \hat{y}_i$$

Residual



Overview

Topics	Type	Difficulty level	Recommended Approach
Basics of Statistics	Theory	Low	Can skip, if you know basic Statistics
Getting started with R	Practical	Medium	Can Skip, If you know R
Introduction to ML	Theory	Low	Can Fast Forward
Data Preprocessing	Theory + Practical	Medium	Must watch
Linear Regression modelling	Theory + Practical	Medium-High	Must watch



Overview

Topics	Type	Difficulty level	Recommended Approach
Basics of Statistics	Theory	Low	Can skip, if you know basic Statistics
Getting started with Python	Practical	Medium	Can Skip, If you know Python
Introduction to ML	Theory	Low	Can Fast Forward
Data Preprocessing	Theory + Practical	Medium	Must watch
Linear Regression modelling	Theory + Practical	Medium-High	Must watch

