Agenda

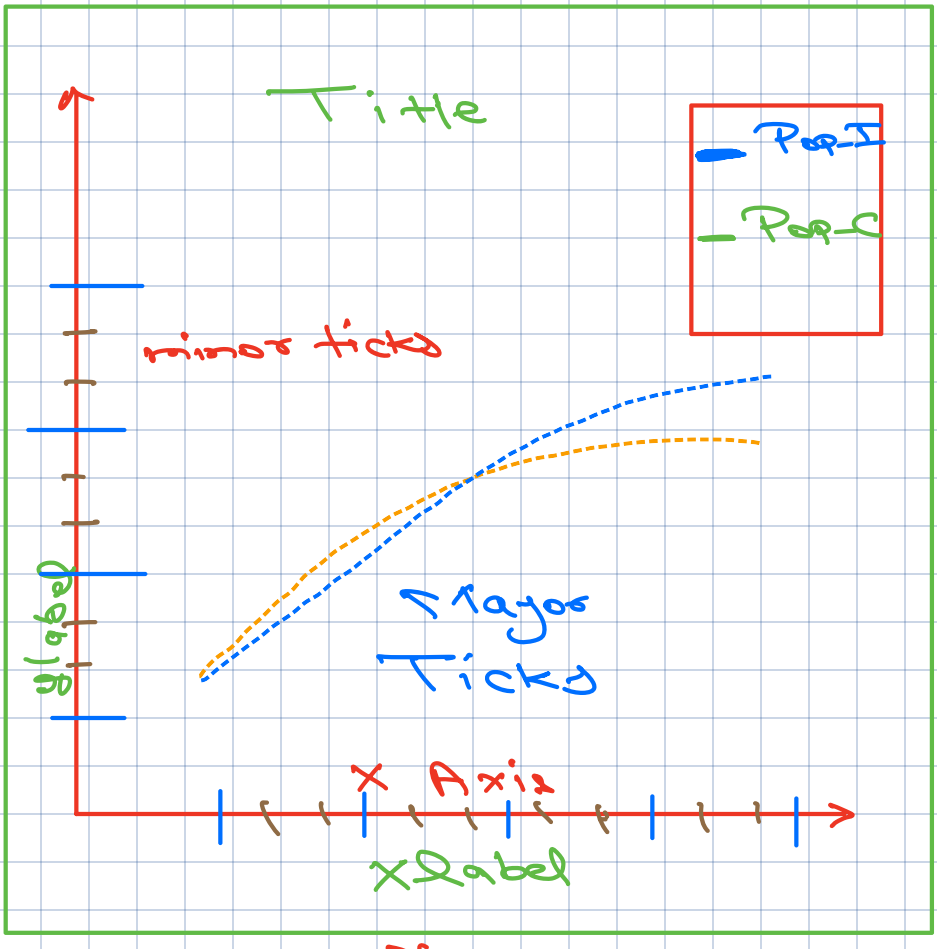
- Intro to Matplotlib and Seaborn
- Tencent Dataset
- Anatomy of a plot
- Univariate Analysis
- Bivariate Analysis

Data
- Rows / Record / DP
- Features / Columns

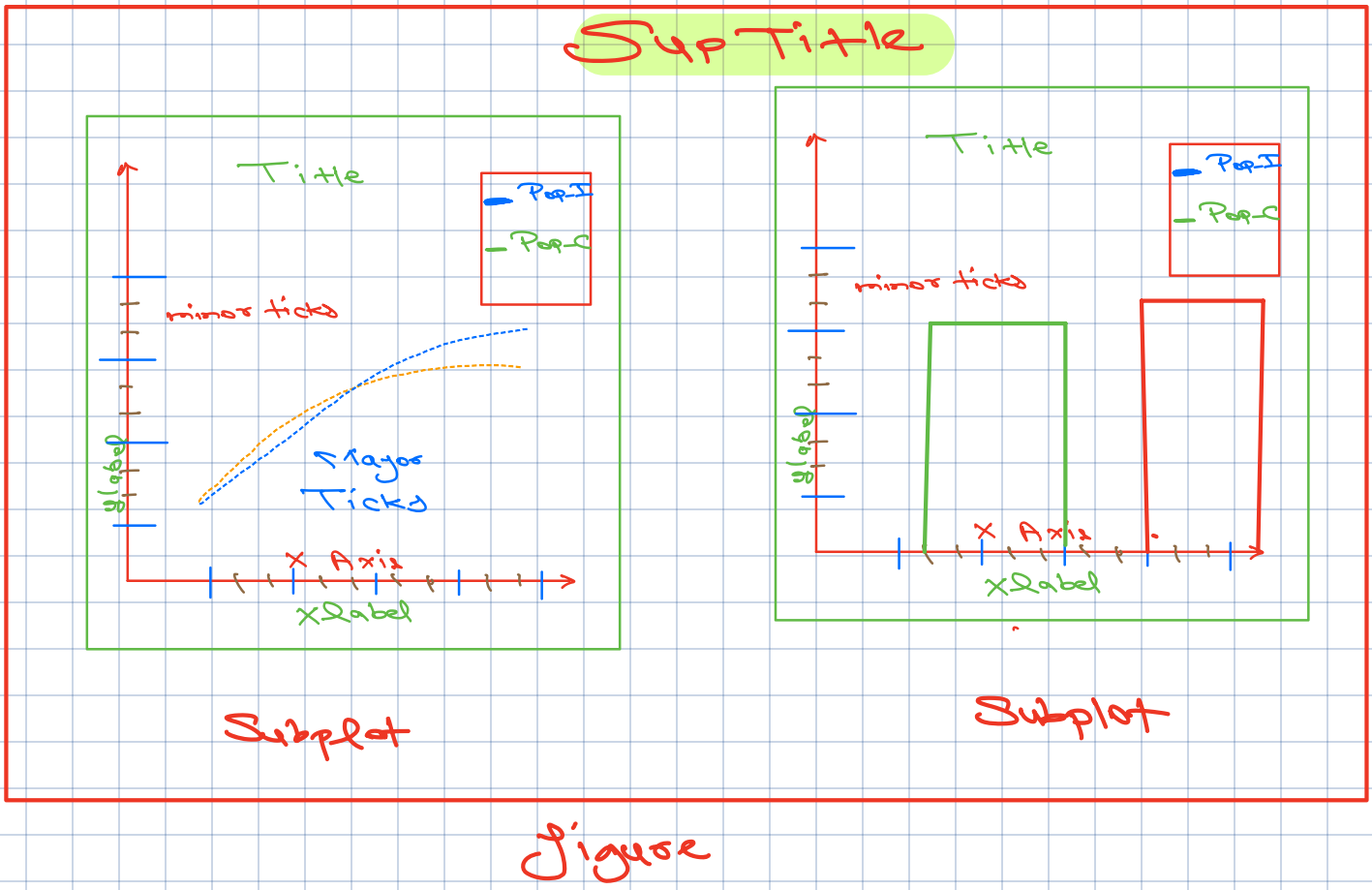
Type of Features / Columns

- Numerical → Bool, int, float ---
- Categorical → Objects / numerical Discrete
  - Ordinal
  - Nominal

- Ordinal → Order Matters
- Nominal → Order doesn't Matter
  Ex Gender



Figure



* Univariate Analysis
  - Only One Variable is Involved

* Bi Variate Analysis
  - Two variables analysed together
    - Relationship

* Multi-Variate Analy
  - ≥ 3 Variable Involved

Univariate Analysis

- Only One Variable in Inve

- Numerical
- Categorical
  - Genre
  - PlatForm
  - ,,
  - ,,

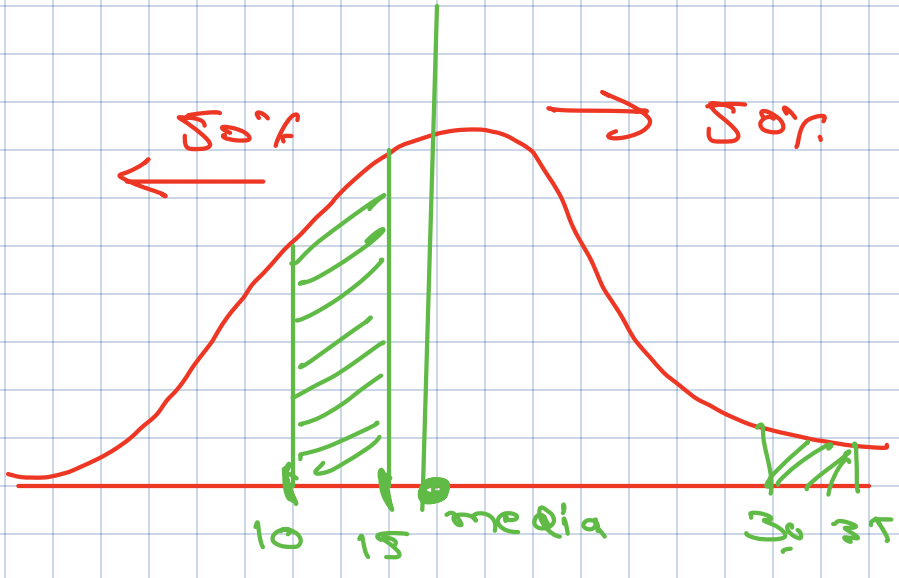
* Genre ⟶ Data

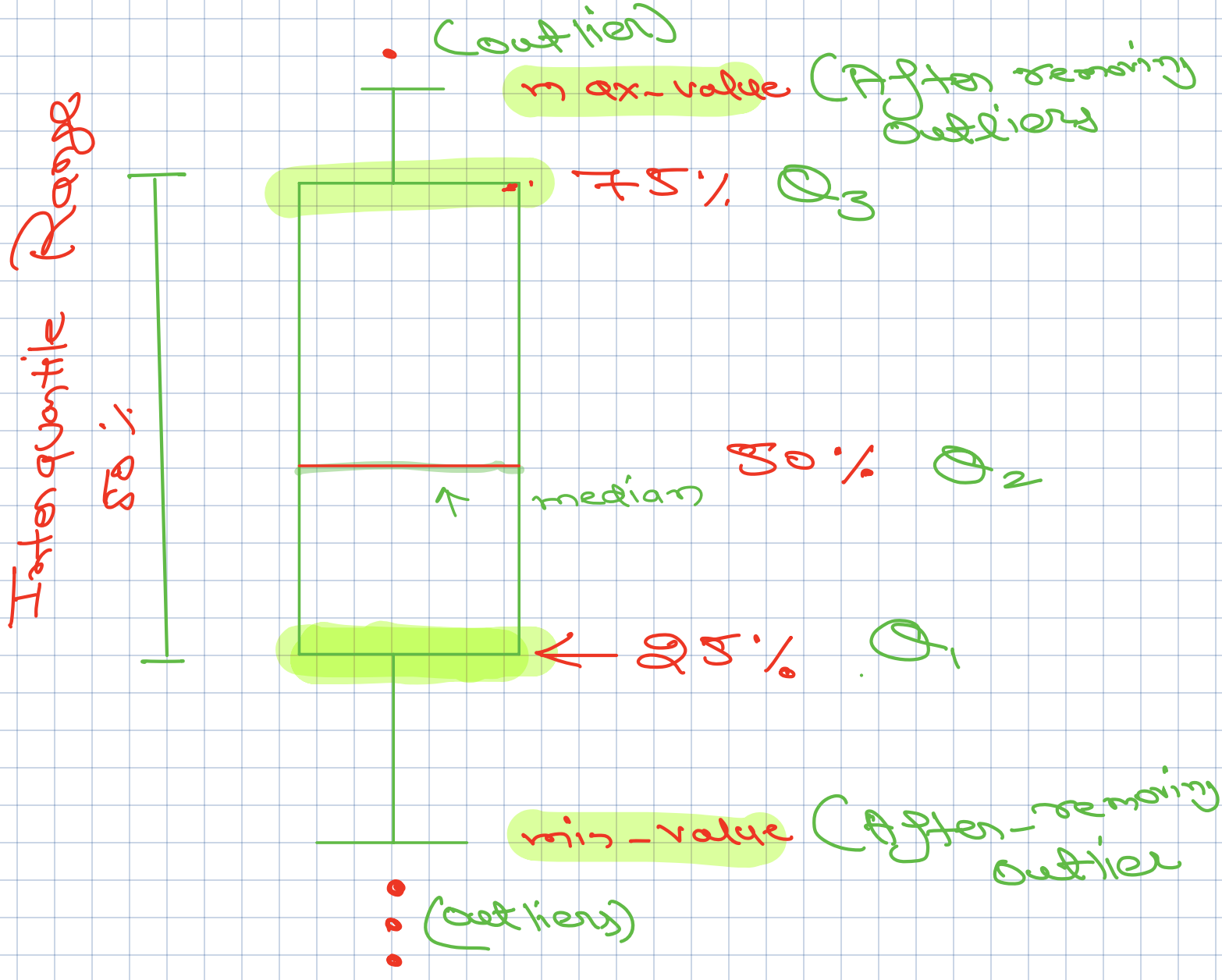
Categorse ⟶ Counts

1. Barchart → Absolute Value
2. Pie Chart → Proportions / Percentages

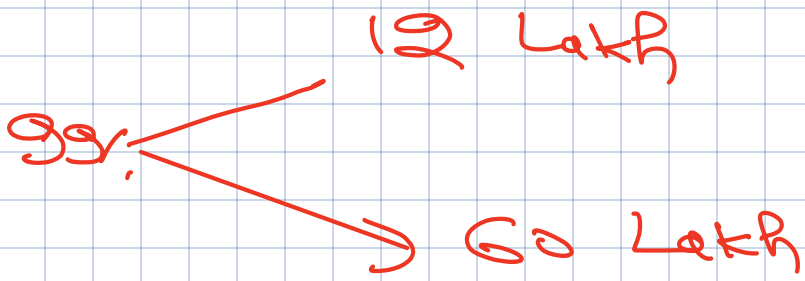
* Histogram
  - Distribution
  - Outliers
  - Skewed / Symmetric
  - Suitable for Numerical Univariates

Goal: H want to present
  - median
  - Min
  - Max
  - Percentile
  - Outliers

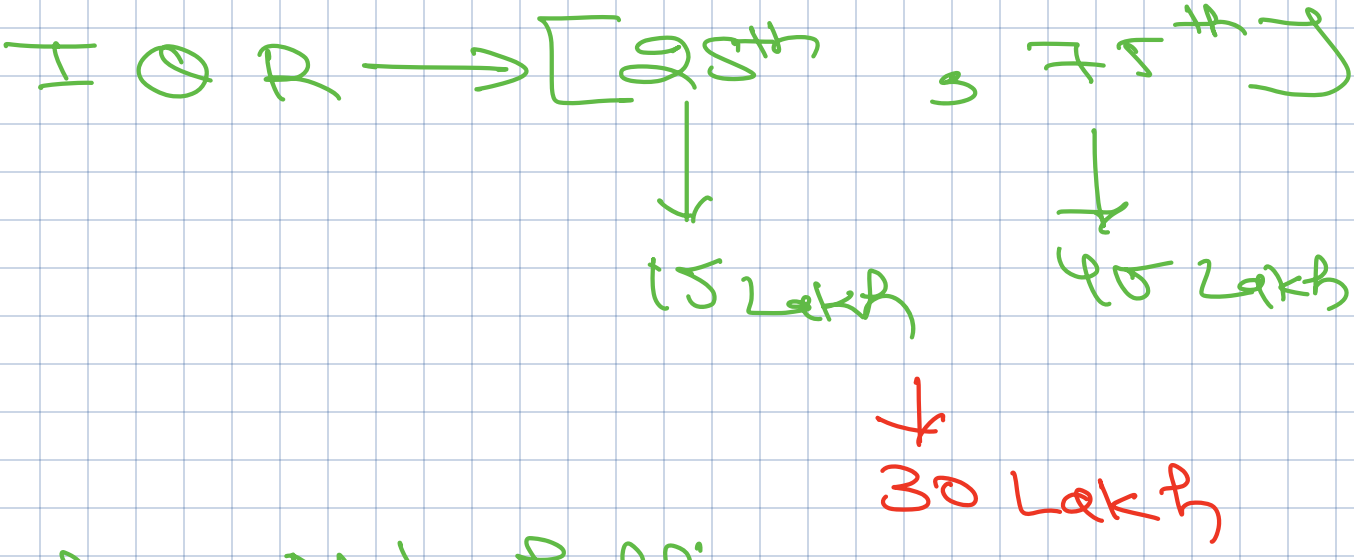




* Outlier → Excessively High or Low Values



10	Cores	Outlier



Values Not Falling in Below Range

$$Q_1 - 1.5\,IQR$$

$$Q_3 + 1.5\,IQR$$

Outliers

IQR = 30

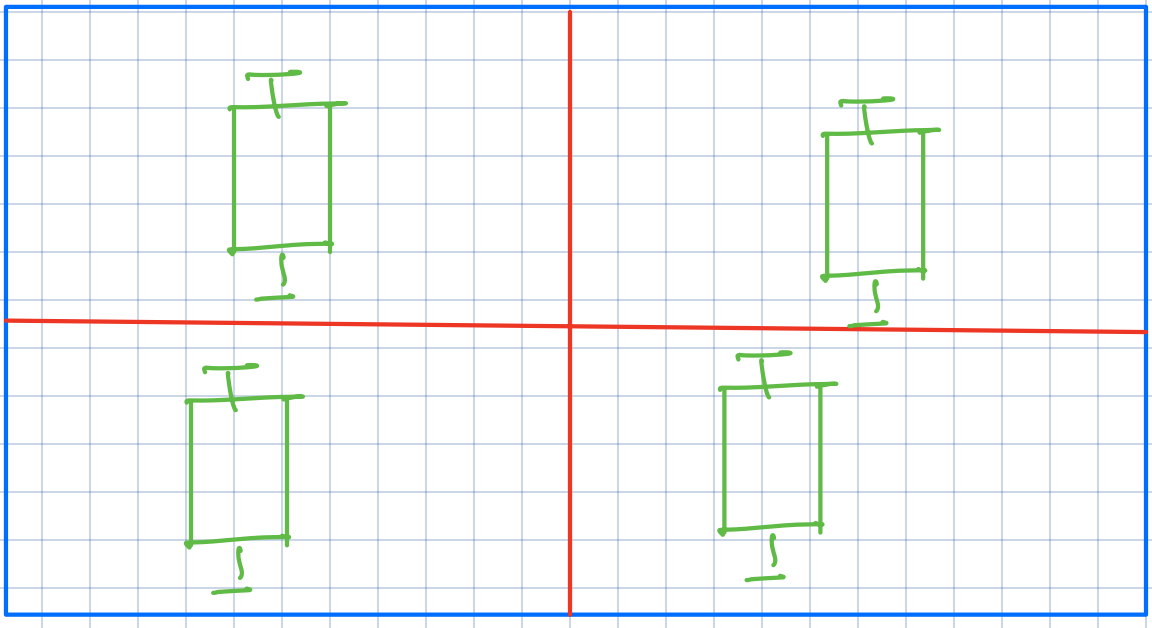
$Q_1 = 15$

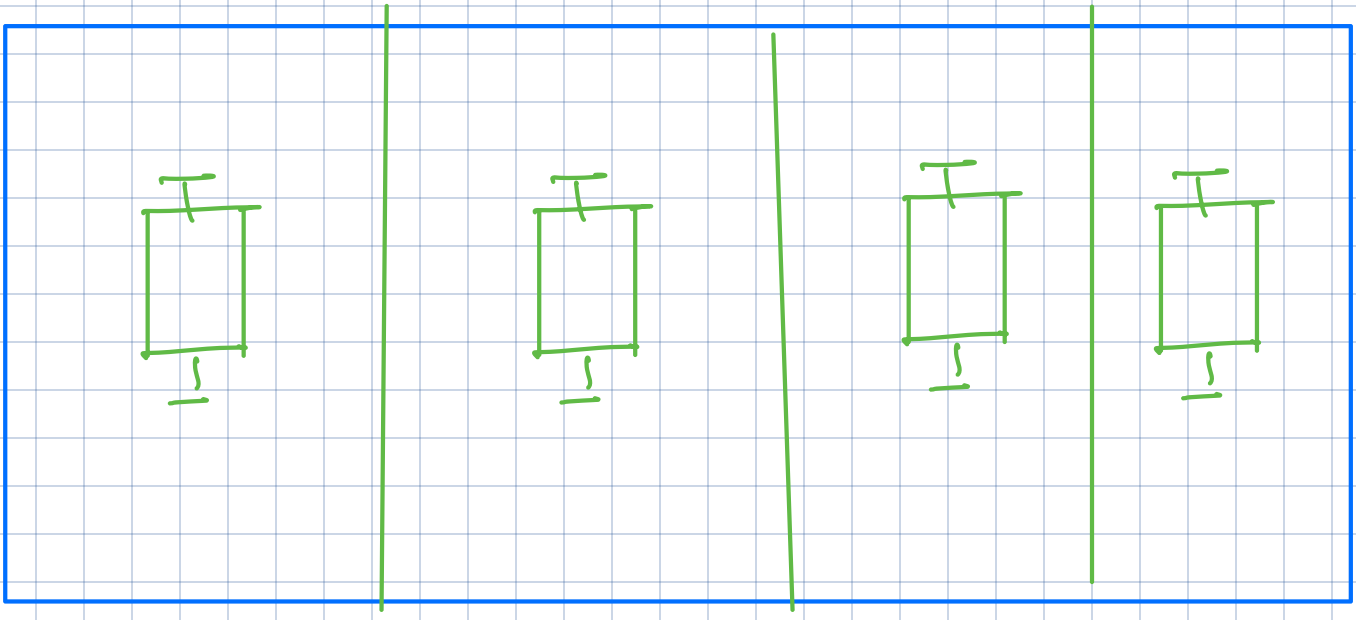
$Q_3 = 45$

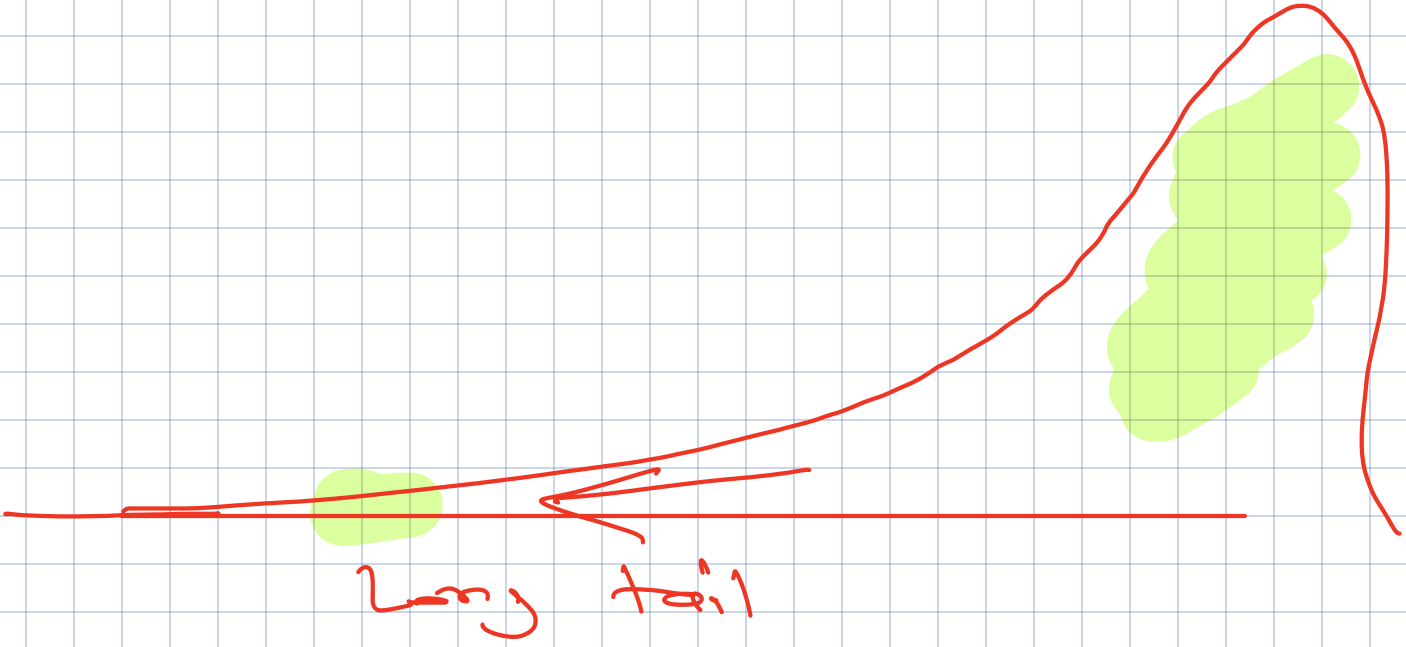
$(15 - 30 \times 1.5 , 45 + 30 \times 1.5)$

$= (-30, 90)$

Ho






Long tail

25 %

33 mark

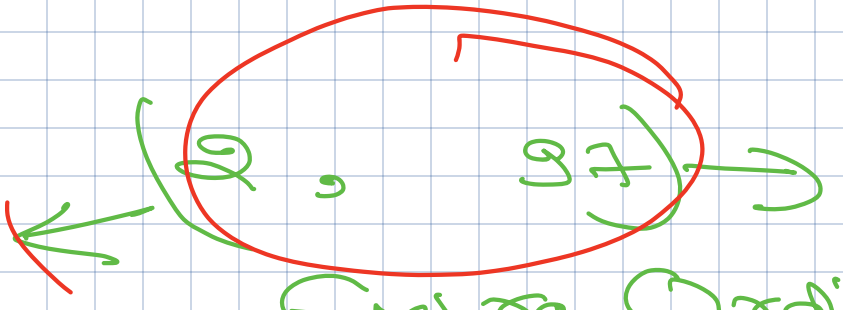
75 %

80 mark

80 - 33 = IQR

3.0

Outlier (33 - 1.5 x IQR, 80 + 1.5 x IQR)



Extra Ordinary