

Topic: Exploratory Data Analysis (EDA)

Presentation of Bivariate Data

Part D: Time Series

School of Mathematics and Applied Statistics



Two Continuous Variables with one as Time: Where in the statistical process?

- Ethics
- **Nature of the question to be answered**
 - ✧ *What is the nature of variation over time?*
- Context/Expertise
- Design:
 - Experiment vs. observational study
 - Sampling
 - Measurement
- **Description and analysis:**
 - ✓ *Line plots and Scatterplots*
- Conclusions and decision making

VARIATION

Time Series: How do data vary over time?

A time series is a collection of univariate data in which the values are recorded at successive time intervals

- measurements may be discrete or continuous
- observations may be observed in discrete or continuous time
- **Examples:**
 - Temperature; Rainfall
 - Sales
 - Road deaths by month
 - Share prices
 - Employment
 - Tourist arrivals ...

What might we be interested in when we look at measurements collected over successive time intervals?

An example: BHP share prices

Daily opening share price from 1st Sept to 31st Dec 2015 (n=85 obs)



Sequence	Date	Open Price	Sequence	Date	Open Price
168	1-Sep-2015	34.42	168	1-Sep-2015	34.42
169	2-Sep-2015	35.27	169	2-Sep-2015	35.27
170	3-Sep-2015	34.94	170	3-Sep-2015	34.94
171	4-Sep-2015	33.75	171	4-Sep-2015	33.75
172	8-Sep-2015	34.94	172	8-Sep-2015	34.94
173	9-Sep-2015	35.36	173	9-Sep-2015	35.36
174	10-Sep-2015	34.04	174	10-Sep-2015	34.04
175	11-Sep-2015	34.1	175	11-Sep-2015	34.1
176	14-Sep-2015	33.88	176	14-Sep-2015	33.88
177	15-Sep-2015	33.66	177	15-Sep-2015	33.66
178	16-Sep-2015	34.93	178	16-Sep-2015	34.93
179	17-Sep-2015	35.35	179	17-Sep-2015	35.35
180	18-Sep-2015	34.78	180	18-Sep-2015	34.78
181	21-Sep-2015	34.3	181	21-Sep-2015	34.3
...

- What do these data reveal?

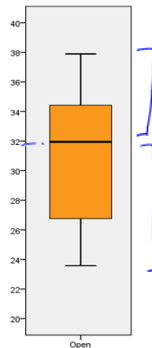
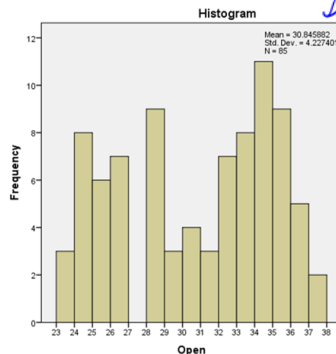
BHP Share Prices: Univariate distribution

What do the univariate plots reveal? no order considered

BHP Open Price Stem-and-Leaf Plot

Frequency	Stem &	Leaf
3.00	2 .	333
14.00	2 .	44444444555555
7.00	2 .	6666666
12.00	2 .	888888889999
7.00	3 .	0000111
15.00	3 .	222222233333333
20.00	3 .	44444444444455555555
7.00	3 .	6666677

Stem width: 10.00000
Each leaf: 1 case(s)



- Shape: Slightly skewed to left
- Centre: median about \$32 and mean is \$30.85
- Spread: range is $\approx 37 - 23 = 15$ and sd is \$4.23

- Outliers: None
- Patterns: None

Time Series

In simple univariate analysis, use

- Stem-and-leaf plots,
- Boxplots
- histograms

to find centre, shape of the distribution, spread, outliers, patterns (S&L)

But these plots ignore the **time dimension**

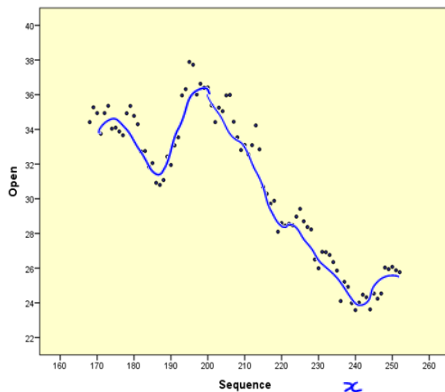
So if there is a series of data points over time, we need a **different technique** to reveal the components of the time series

To represent a **time series** a line plot is more appropriate.

BHP Share Prices: Time Series

Scatterplot: Opening share price on vertical axis against time on the horizontal axis.

What does it reveal?



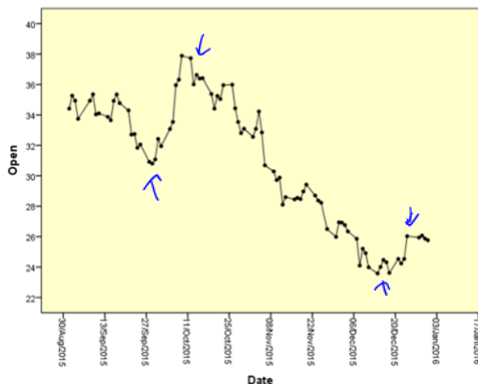
- There is quite a bit of **variation or scatter**
- But there is an **overall pattern** of price over time
- This is called the trend

BHP Share Prices: Time Series

Line Plot: Price by time: Join the data points in the scatterplot in sequence order

What does it reveal?

BHP Daily Opening Share Price: 1st Sept - 31st Dec 2015



- There are **fluctuations** within the overall trend

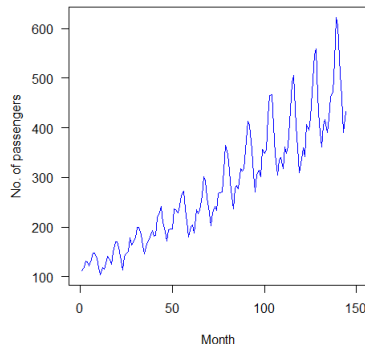
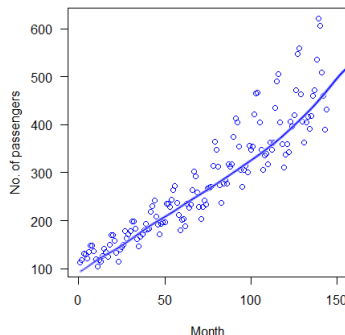
- We can see highs (peaks) and lows (troughs)

Components of a Time Series

A time series may consist of components such as

- a trend
- cyclical and /or seasonal variation
- random variation

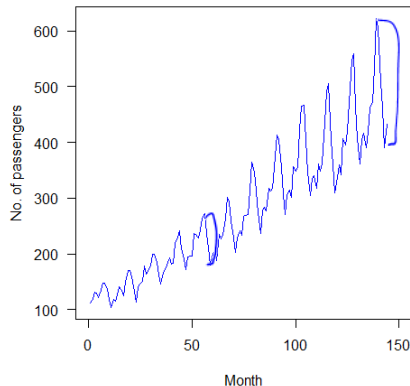
Example: Number of airline passengers over time



1949 → 1960 Jan - Dec
⇒ 144 data points

Seasonal Patterns

Example: Number of airline passengers over time

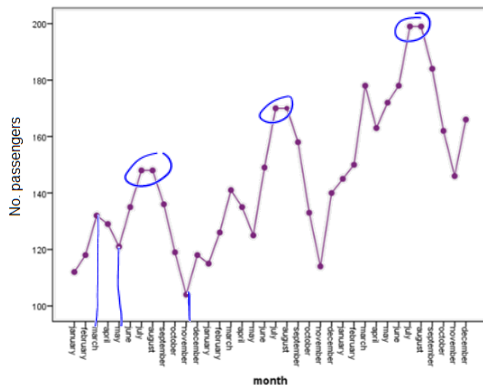


What do you observe?

- increase in number of passengers over time.
- Appears to be a seasonal pattern
- Peaks within the repeating pattern are getting higher over time
- More variation in later years
- This is called a **multiplicative time series**

Seasonal Patterns - Zooming in

Number of airline passengers over time : 3-year slice



What else can you observe?

- Appears to be a seasonal pattern
 - First peak each year at about March
 - Dip at about May
 - High peak at about Jul-Aug
 - Low occurs about Nov., 3 months after high peak
- Not entirely same pattern each year

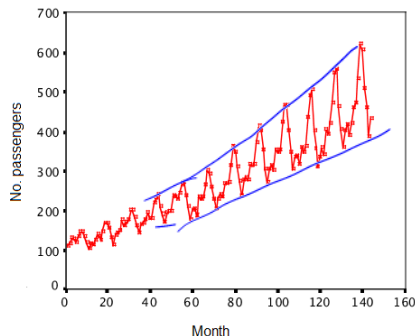
Transformed data: $\log(\text{passengers})$

Example: Transform by taking the log of Number of airline passengers over time

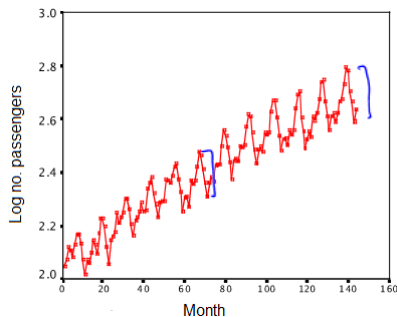
What is the impact of the transformation?

- Removes the multiplicative behaviour of the seasonal pattern
- This results in an additive time series

Original



Transformed



Dynamic Plots Example: CBA ASX Chart

You may see dynamic interactive plots such as this one (see link below): customize the plot by choosing the time period, labelling a particular point, adding a moving average, comparing to another series.

Commonwealth Bank of Australia Shares



Ref: <https://www.marketindex.com.au/asx/cba>

Dynamic plots - try these

Go to:

<https://ourworldindata.org/covid-vaccinations>

Share of people who received at least one dose of COVID-19 vaccine -
- go to website - play videos

- Select countries
- CHART
- MAP

Also go to:

<https://ourworldindata.org/>

to find plots on other interesting topics such as Artificial Intelligence.