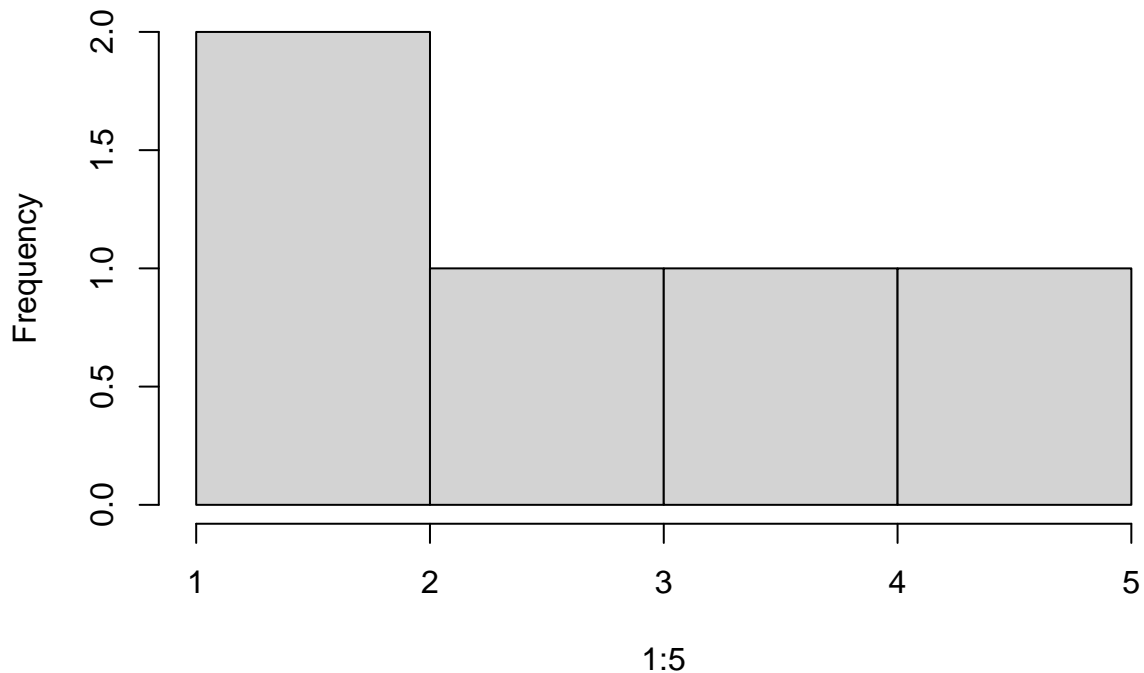# Tutorial 1 RMD

Ze Yu

Tutorial 1 RMD

This first tutorial will go through how the elementary basics of getting started with R with a focus on high dimensional time series modelling.

R Basics

Histogram of 1:5

Functions and Variables

Packages

Useful Package "EcoSystems"

Coding without the help of packages is also known as coding in "base" R. Although there may be legitimate reasons for doing this (such as building your own package and not reducing dependencies), it is not recommended for practical purposes.

Some packages are very large and involve many different packages.

The pre-eminent example of this is the aforementioned tidyverse package, which is an example of a package "ecosystem".

Other ecosystems exist.

For the purposes of high dimensional time series regression, there are broadly two other ecosystems you should be aware of and which we will be using.

Tidyverse

forecast

The forecast package is developed mainly for use with ts data structures. The maintainers of the package have introduced a new tidy time series data structure, and as such most of the functionality of forecast has been superseded byt he "tidyverts" ecosystem which uses the tsibble data structure.

Tidy time series (tidyverts)

The tidyverts ecosystem consists of tsibble, fable (forecasting), and feats (feature extraction).

zoo/xts ecosystem

Technically not an ecosystem by themselves, zoo and xts are extensions of the base ts data structure to multiple time series.

zoo is an extension of ts, and xts is in turn an extension of zoo.

In general, xts should be be backwards compatible with zoo and in turn ts, however this is not always the case.

Many packages such as strucchange (for structural breaks), rugarch (garch models), rmgarch, stochvol, etc work with built in support for zoo formats, but not tsibble.

These data structures are much older than tsibble, and as such for many older packages will actually have much better comptability.

tsbox

Manually converting between the different time series formats can be very tedious and prone to errors.

In lieue of the (in)famous xkcd, tsbox provides a selection of functions which will seamlessly convert different formats into one another.

It also introduces a tsbox() class which is meant to be class agnostic (converted on the fyl as necessary), but this is finicky to get working and not recommdned.

As a bonus, it also contains a selection of very useful plotting functions.

Data Structures

The previous section provided a run down of the different ecosystems available. Which one should you use? The answer is like any good analyst, all of them. Similar to how there is no one answer for, say, should one learn R/Python/Julia, there is no one answer, and often it is best to work with a mix of all options.

In general, tsibble is best for exploratory data analysis (cleaning, plotting), whereas zoo/xts have much wider support for modelling via packages.

As such, it is important that you are comfortable with seamlessly working with all different data structures, understand what limitations they all have and hence when to use one over another.

Data Strucutres for Time series

Converting between different Time Series Formats

ARIMA Modelling

We end this tutorial by finishing with some basic ARIMA modelling There are multiple different implementations available for fitting ARIMA models.

Tutorial 2 - EDA of Time Series ————————————————————

This tutorial will focus on Exploratory Data Analysis (EDA) of high dimensional time series.

Exploratory Data Analysis is the most important step in any project pipeline. This is the first step, and any errors you make here will show up and affect all subsequent results.

Unfortunately this is also the most boring step, and can be quite tedious.

Exploratory Data Analysis encompasses data collection, data cleaning, descriptive analysis.

Although this may seem simple, this step alone is so large that there are entire careers that can be spawned out of this.

```
library(tidyverse)
library(feasts)
library(tsbox)
```

Missing Data

Imputation

Descriptive Analysis

Data Visualization

Data visualization is typically a course onto itself.

The line plot is the most basic (t on x axis, y variable on y axis), and of course you can extend this to plot multiple different time series. A good rule of thumb for visualizing multiple time series to plot, at most, 4 time series. This of course depends on the specific dataset, but generally keeps the resulting line plots readable and not too messy.

Time Series Features

The most technically demanding aspect of this tutorial is the construction of different time series features and their subsequent analysis.

Which features you

Tutorial 3 -

Tutorial 4 -

Tutorial 5 -

Tutorial 6 -

Tutorial 7 -

Tutorial 8 -

Tutorial 9 -

Tutorial 10 - Dynamic Factor Models

Tutorial 11 -

Tutorial 12 -