

ETF3231/5231: Business forecasting

Week 1: Intro to forecasting and R
<https://bf.numbat.space/>



Lecturer: Professor George Athanasopoulos

■ Contact details

- ▶ Room H5.83, Building H, Caulfield.
- ▶ Consultation: Thursday 3.30-4.30pm (subject to changes).
- ▶ All general discussion questions will be answered on the discussion forum: <https://edstem.org/au/courses/16026> (check for answers before you ask).
- ▶ Assignment consultations - see your tutors.

Lecturer: Professor George Athanasopoulos

■ Contact details

- ▶ Room H5.83, Building H, Caulfield.
- ▶ Consultation: Thursday 3.30-4.30pm (subject to changes).
- ▶ All general discussion questions will be answered on the discussion forum: <https://edstem.org/au/courses/16026> (check for answers before you ask).
I will not consult via email.
- ▶ Assignment consultations - see your tutors.

Lecturer: Professor George Athanasopoulos

■ Contact details

- ▶ Room H5.83, Building H, Caulfield.
- ▶ Consultation: Thursday 3.30-4.30pm (subject to changes).
- ▶ All general discussion questions will be answered on the discussion forum: <https://edstem.org/au/courses/16026> (check for answers before you ask). I will not consult via email.
- ▶ Assignment consultations - see your tutors.

■ Seminars (in person, 2pm every Thursday, Room K321)

■ Tutorials

- ▶ In person for all students.

- Nadeeka Dilini Basnayake, Monday 5-6.30pm, zoom.
- Ari Handayani, Tuesday, 11-12.30pm, zoom.
- Kulan Ranasinghe, Wednesday, 12-1pm, 3-3.30pm. Room H4.62.
- George Athanasopoulos, Thursday, 3.30-4.30pm, Room H5.83.
- Joan Tan (Head Tutor), Friday 12-1pm, zoom.

Speak to your tutor if you would prefer a face-to-face consultation to see whether that can be arranged.

Brief bio: George Athanasopoulos

- Professor and Head of Department of Econometrics and Business Statistics, Monash Business School.
- President: *International Institute Forecasters*
 - ▶ Bridge the gap between theory and practice, with practice helping to set the research agenda and research providing useful results.
- Associate Editor: *International Journal Forecasting*
 - ▶ The leading academic journal in business forecasting.
- Editorial board: *Journal of Travel Research*

How my forecasting methodology is used:

- Forecasting Australian retail sector
- Australian tourism (latest is post-Covid19)
- Hospital admissions (UK and Mornington Peninsula)
- Monash student enrollment numbers
- Australian prison populations
- Macroeconomic variables
- Restaurant bookings

Unit objectives

- Obtain an understanding of common statistical methods used in business and economic forecasting.
- Learn how to build accurate and robust models for forecasting.
- Acquire computer skills vital for forecasting business and economic data.
- To gain insights into the problems of implementing and operating large scale forecasting systems for use in business.

We'll use R to do all this - so the course is about learning good forecasting practices using a very powerful tool.

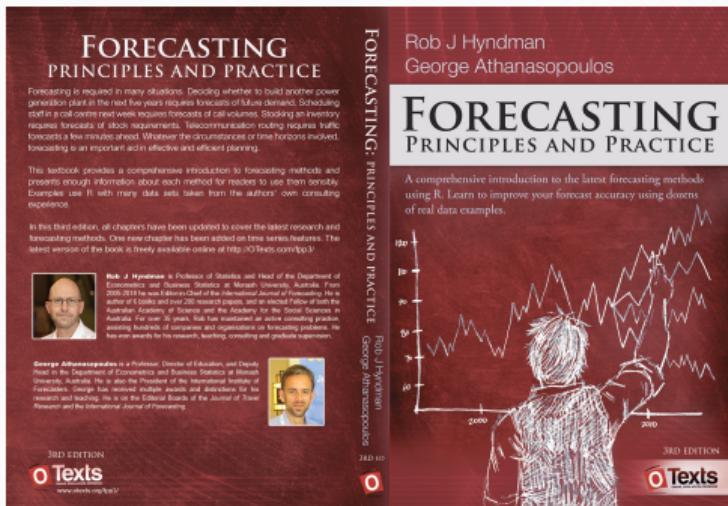
Teaching and learning approach

- **Pre-seminar preparation:** watch recorded lectures embedded in the textbook at <http://OTexts.org/fpp3/> and read the book sections. Min 60 minutes to do this.
- **One 50 minute in person seminar.** This will be mostly an interactive workshop style. The plan is that I present a quick summary of the material and then go through examples with coding (you will also get a chance to practice with me) or deeper explanations or proofs when required. Hence, it is important that you are ready to go, with R installed, on your laptops during the lecture every week.
- **One 80 minute computer lab session.** Try and prepare questions before you go to the lab (a much more enjoyable experience).

Textbook - key reference

Hyndman, R. J. & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd Edn.

- <http://OTexts.org/fpp3/>
- Free online
- Printed version available [here](#)
- Data sets in associated package.
- R code for examples



Software



Available for download from CRAN:

<https://cran.r-project.org>



Available for download from RStudio:

<https://www.rstudio.com/products/rstudio/download/>

Software



<https://PollEv.com/georgeathana023>

How familiar are you with R, RStudio?

Available for download from CRAN:

<https://cran.r-project.org>



Available for download from RStudio:

<https://www.rstudio.com/products/rstudio/download/>

Main packages



Main packages

```
# Install required packages (do once)
install.packages(c("tidyverse", "fpp3", "GGally"))
```

Main packages

```
# Install required packages (do once)
install.packages(c("tidyverse", "fpp3", "GGally"))
```

```
# At the start of each session
library(fpp3)
```

Main packages

```
# Install required packages (do once)
install.packages(c("tidyverse", "fpp3", "GGally"))
```

```
# At the start of each session
library(fpp3)
```

```
# Data manipulation and plotting functions
library(tidyverse)
# Time series manipulation
library(tsibble)
# Tidy time series data
library(tsibbledata)
# Time series graphics and statistics
library(feasts)
# Forecasting functions
library(fable)
```

Week 1 homework

Pre-class tutorial activity

- Install R, RStudio and required packages
- See <https://otexts.com/fpp3/appendix-using-r.html>
- `install.packages(c("tidyverse", "fpp3", "GGally"), dependencies = TRUE)`

Complete pre- or in-class

- Complete chapters 1-5, 8 in <https://learnr.numbat.space/>

Reading and Lecture Videos

- Chapter 1 of <https://otexts.com/fpp3>

Week 1 homework

Pre-class tutorial activity

- Install R, RStudio and required packages
- See <https://otexts.com/fpp3/appendix-using-r.html>
- `install.packages(c("tidyverse", "fpp3", "GGally"), dependencies = TRUE)`

Complete pre- or in-class

- Complete chapters 1-5, 8 in <https://learnr.numbat.space/>

Reading and Lecture Videos

- Chapter 1 of <https://otexts.com/fpp3>

Aim: become familiar with R, RStudio and the tidyverse packages

Some other references

■ Forecasting textbooks

- ▶ Ord, J. K., Fildes, R. & Kourentzes, N. (2017). *Principles of Business Forecasting*. 2nd ed. Wessex Press Publishing Co.
- ▶ Gloria Gonzalez-Rivera (2013) *Forecasting for Economics and Business*, Pearson/Addison-Wesley.

■ R

- ▶ Adler, J. (2012). *R in a Nutshell*. 2nd ed. O'Reilly.
- ▶ ModernDive: Statistical Inference via Data Science (Ismay, C. & Kim, A. Y. 2020). Chapters 1-4. <https://moderndive.netlify.com/>
- ▶ RYouWithMe? R for beginners. Lisa Williams, UNSW, Co-founder of R-Ladies Sydney.
<https://rladiessydney.org/courses/ryouwithme/>

Outline

Week	Topic	Chapter
1	Introduction to forecasting and R	1, App
2	Time series graphics	2
3	Decomposition	3
4	The forecaster's toolbox	5
5–6	Exponential smoothing	8
7–9	Forecasting with ARIMA models	9
10–11	Multiple regression and forecasting	7
11–12	Dynamic regression	10

Assessment

- All students: 4 short individual assignments (IA).
 - ETF5231: extra 4 group assignments (GA) (see next slide).
 - Assignments: total weight 40%
 - Exam (2 hours): weight 60%.
 - Must get at least 45% on exam and 50% overall to pass the unit.
-
- Assignment submission dates are to be confirmed as we go along.
↳ Dates have been corrected.
 - Ass 1 already posted. Will announce shortly.

Assignment schedule

Cohort	Week	Assessment task	Weight
ETF3231+ETF5231	2	IA1	5%
ETF5231	3	GA1	5%
ETF3231+ETF5231	4	IA2	7%
ETF5231	5	GA2	7%
ETF3231+ETF5231	8	IA3	10%
ETF5231	9	GA3	10%
ETF3231+ETF5231	11	IA4	18%
ETF5231	12	GA4	18%

For ETF5231 your mark allocated to assignments will come from individual assignments (weight 0.7) and from group assignments (weight 0.3). E.g. Ass 3 mark will be: $8 \times (0.7) + 5 \times (0.3) = 7.1$

Webpage <https://bf.numbat.space/>

- Includes all lecture note handouts, R code, assignments.
- Forum for asking questions, etc.
- Assignment submissions (connected to moodle page).
- A common question: are the lectures recorded? Yes but... see Unit information in Moodle

R in lectures

- We will be running examples of R code during lecture so it is important to be ready on your laptop during the lecture.
- Hence, I suggest you install R and RStudio straight away on your own laptops.

International Institute of Forecasters Best Student Award



- The IIF provides a prize to the top student in this subject each year.
- A certificate of achievement from the IIF.
- US\$100 plus one year membership.

<https://forecasters.org/programs/research-awards/students>

Assignment 1

Assignment 1

Go to <https://etf3231.github.io/bf/assignments/A1.html>.

Assignment 1: scoring

y = actual, \hat{y} = point forecast, $[\hat{\ell}, \hat{u}]$ = prediction interval

Point forecasts:

$$\text{Absolute Error} = |y - \hat{y}|$$

- Rank results for all students in class
- Add ranks across all five items

Prediction intervals:

$$\text{Interval Score} = (\hat{u} - \hat{\ell}) + 10(\hat{\ell} - y)_+ + 10(y - \hat{u})_+$$

- $u_+ = \max(u, 0)$
- Rank results for all students
- Add ranks across all five items

Chapter 1

What can we forecast?



Which is easiest to forecast?

- 1 Google stock price tomorrow
- 2 Tourism demand next summer
- 3 Google stock price in 6 months time
- 4 Exchange rate of \$US/AUS next week
- 5 Prison population in 2 years
- 6 Time of sunrise this day next year
- 7 Maximum temperature tomorrow
- 8 Daily electricity demand in 3 days time

Which is easiest to forecast?

- 1 Google stock price tomorrow
- 2 Tourism demand next summer
- 3 Google stock price in 6 months time
- 4 Exchange rate of \$US/AUS next week
- 5 Prison population in 2 years
- 6 Time of sunrise this day next year
- 7 Maximum temperature tomorrow
- 8 Daily electricity demand in 3 days time



<https://PollEv.com/georgeathana023>

Which one is easiest to forecast?

Which is easiest to forecast?

- 1 Google stock price in 6 months time
 - 2 Exchange rate of \$US/AUS next week
 - 3 Google stock price tomorrow
 - 4 Prison population in 2 years
 - 5 Tourism demand next summer
 - 6 Daily electricity demand in 3 days time
 - 7 Maximum temperature tomorrow
 - 8 Time of sunrise this day next year
-
- The list is annotated with handwritten text and brackets:
- Items 1, 2, and 3 are grouped together with an orange bracket on the right and labeled "Hard" in orange.
 - Items 4, 5, and 6 are grouped together with a blue bracket on the right and labeled "Medium" in blue.
 - Items 7 and 8 are grouped together with a green bracket on the right and labeled "Easy" in green.

Which is easiest to forecast?

- 1 Google stock price in 6 months time
- 2 Exchange rate of \$US/AUS next week
- 3 Google stock price tomorrow
- 4 Prison population in 2 years
- 5 Tourism demand next summer
- 6 Daily electricity demand in 3 days time
- 7 Maximum temperature tomorrow
- 8 Time of sunrise this day next year

- how do we measure “easiest”?
- what makes something easy/difficult to forecast?

Factors affecting forecastability

Something is easier to forecast if:

- 1 we have a good understanding of the factors that contribute to it;
- 2 there is lots of data available;
- 3 the future is somewhat similar to the past; (*Covid 19, GFC 07/08*)
- 4 the forecasts cannot affect the thing we are trying to forecast.

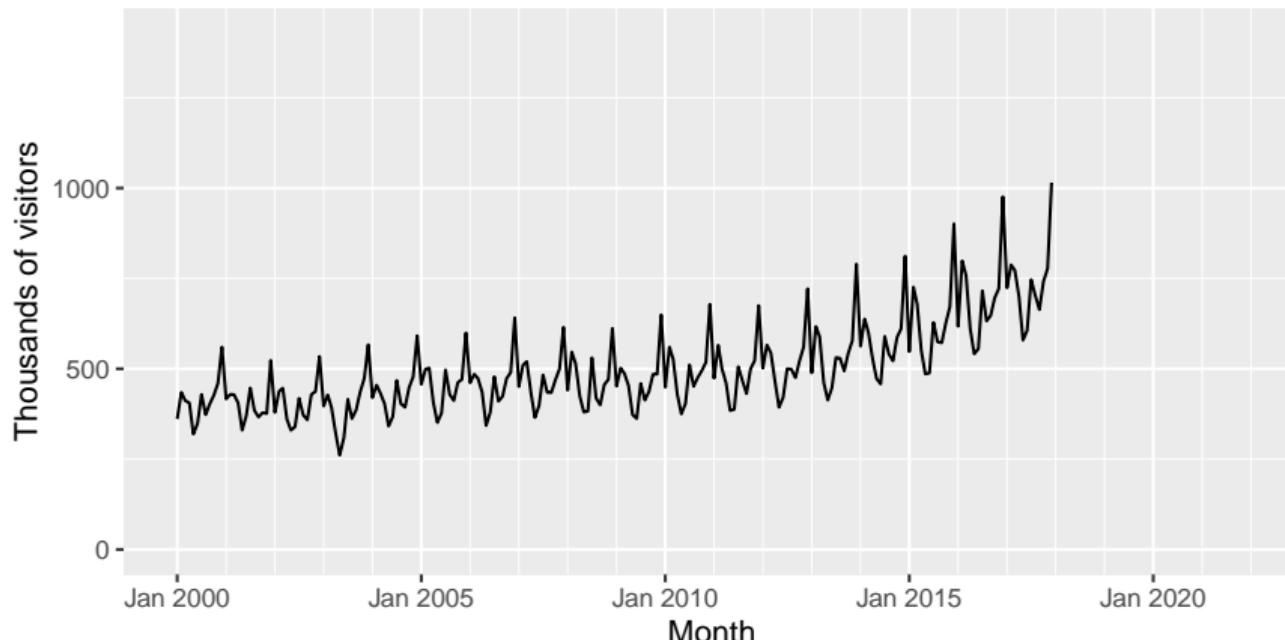
What is a forecast?

A forecast is an estimate of the probabilities of possible futures.

What is a forecast?

A forecast is an estimate of the probabilities of possible futures.

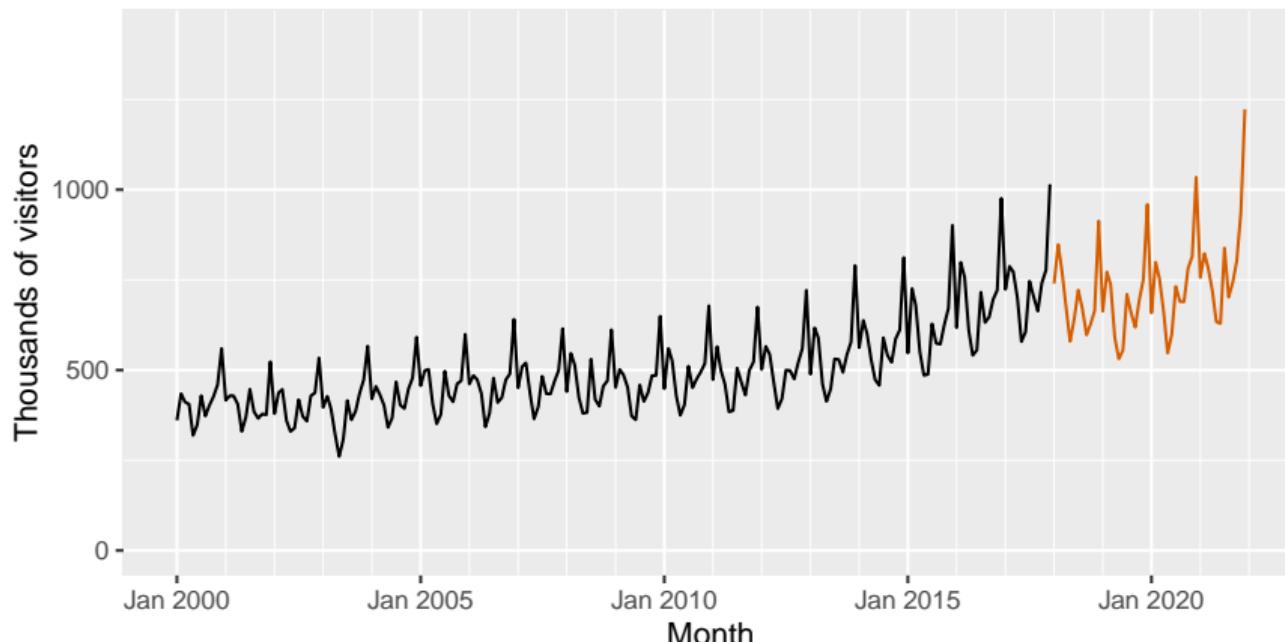
Total short-term visitors to Australia



Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia



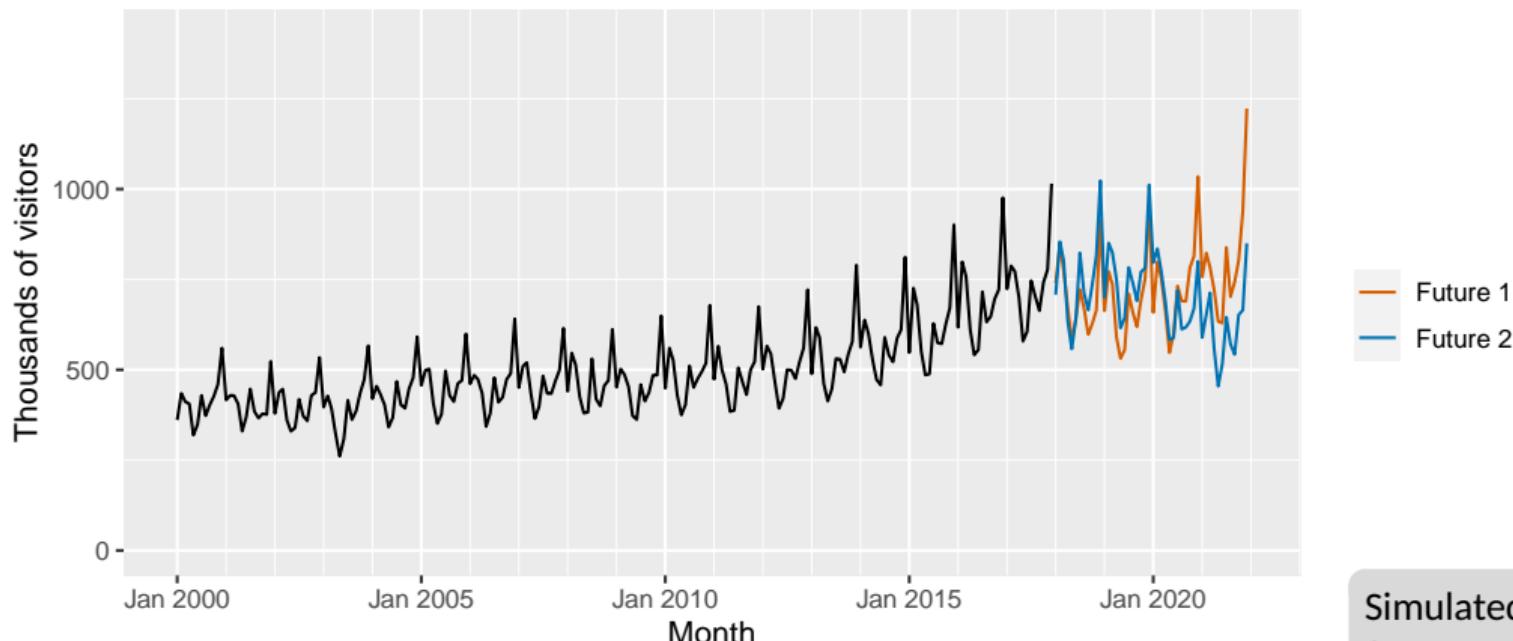
This is not a forecast.

Simulated futures from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia

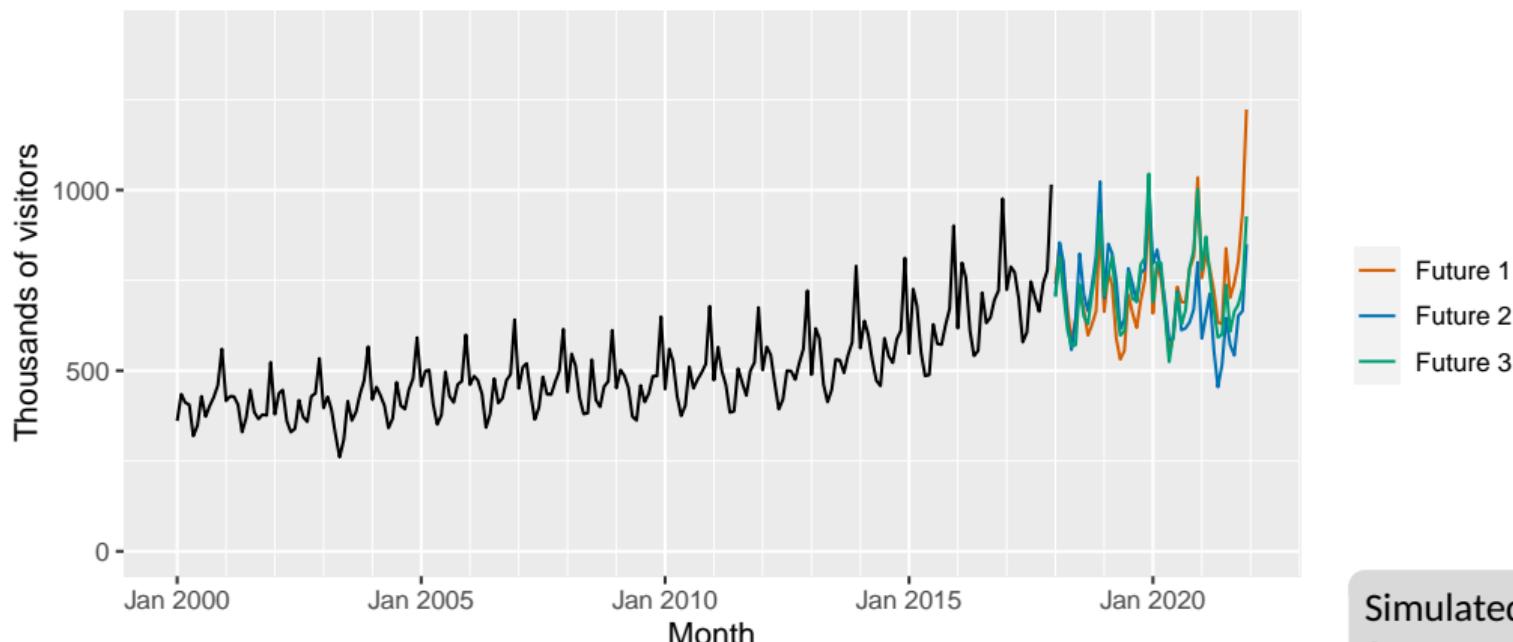


Simulated futures
from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia

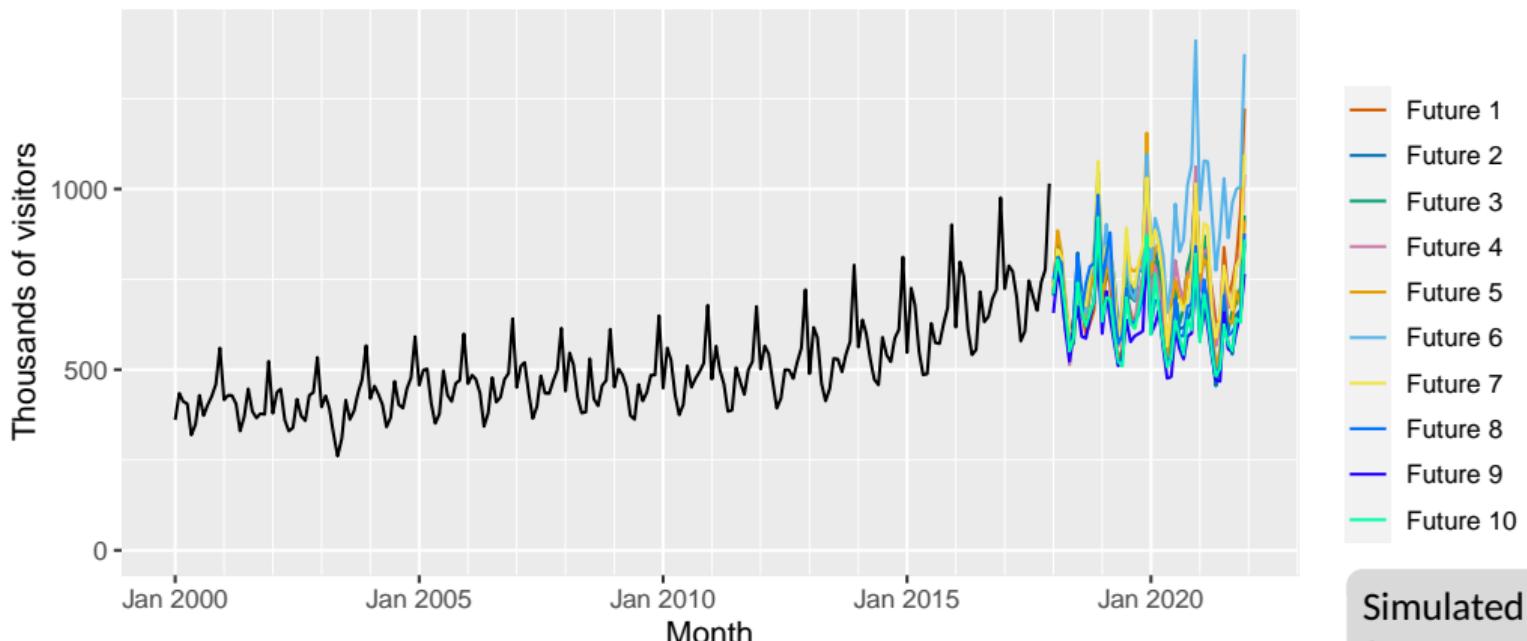


Simulated futures
from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia

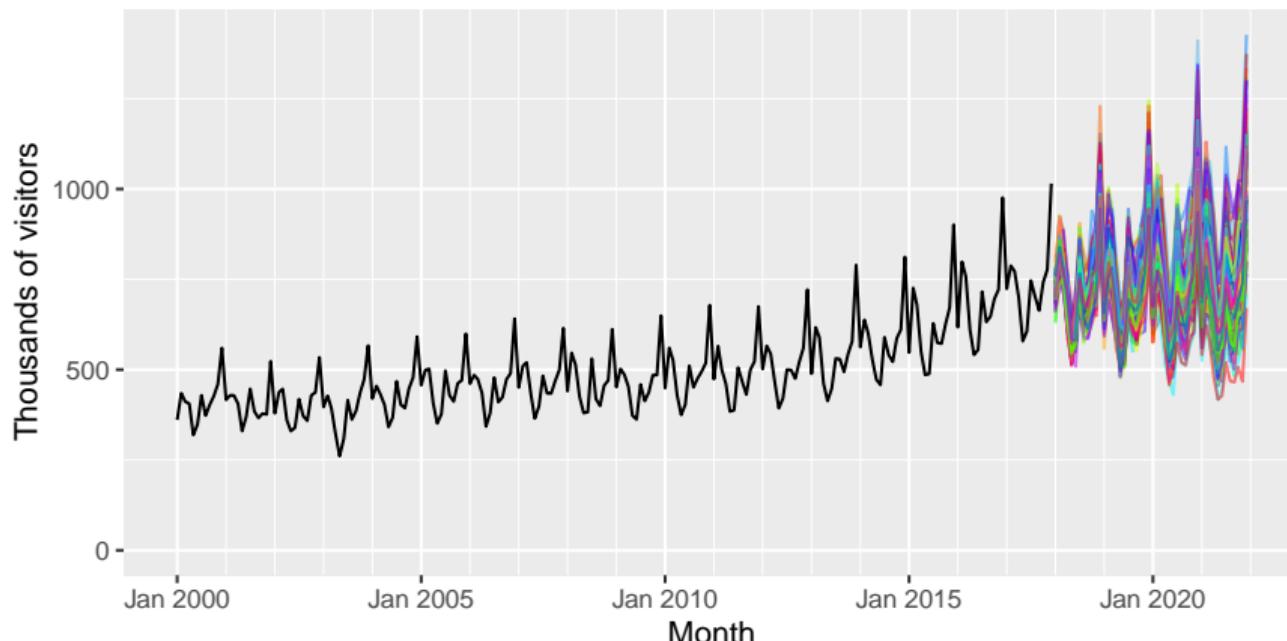


Simulated futures
from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia

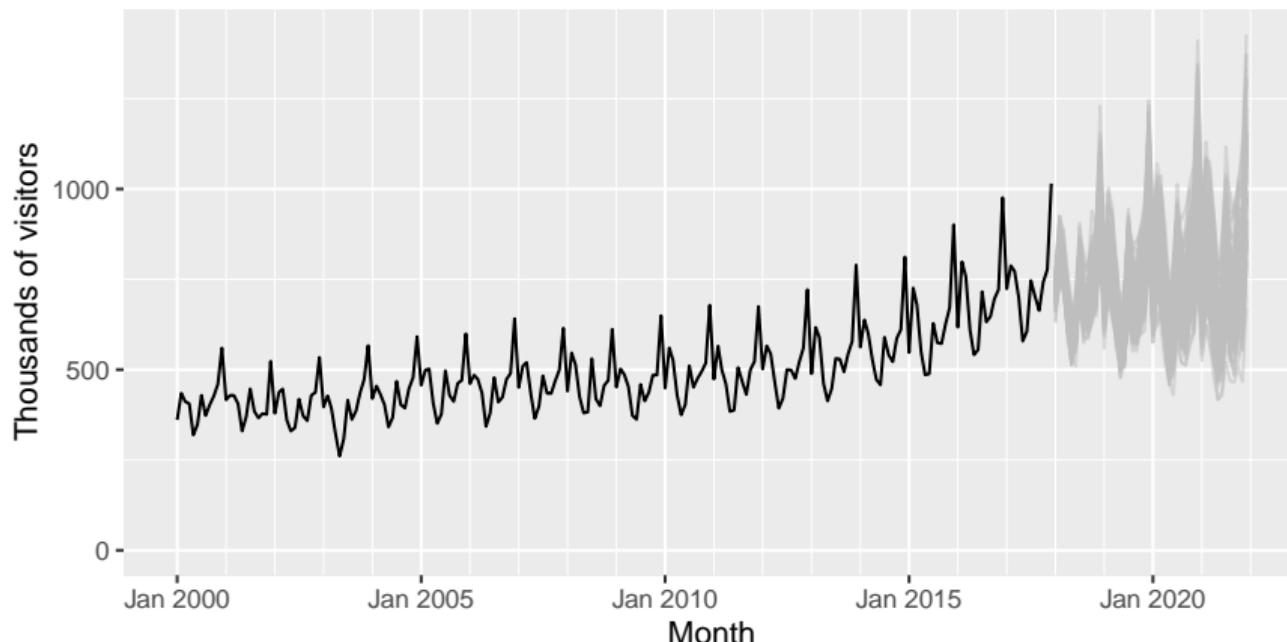


Simulated futures
from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia

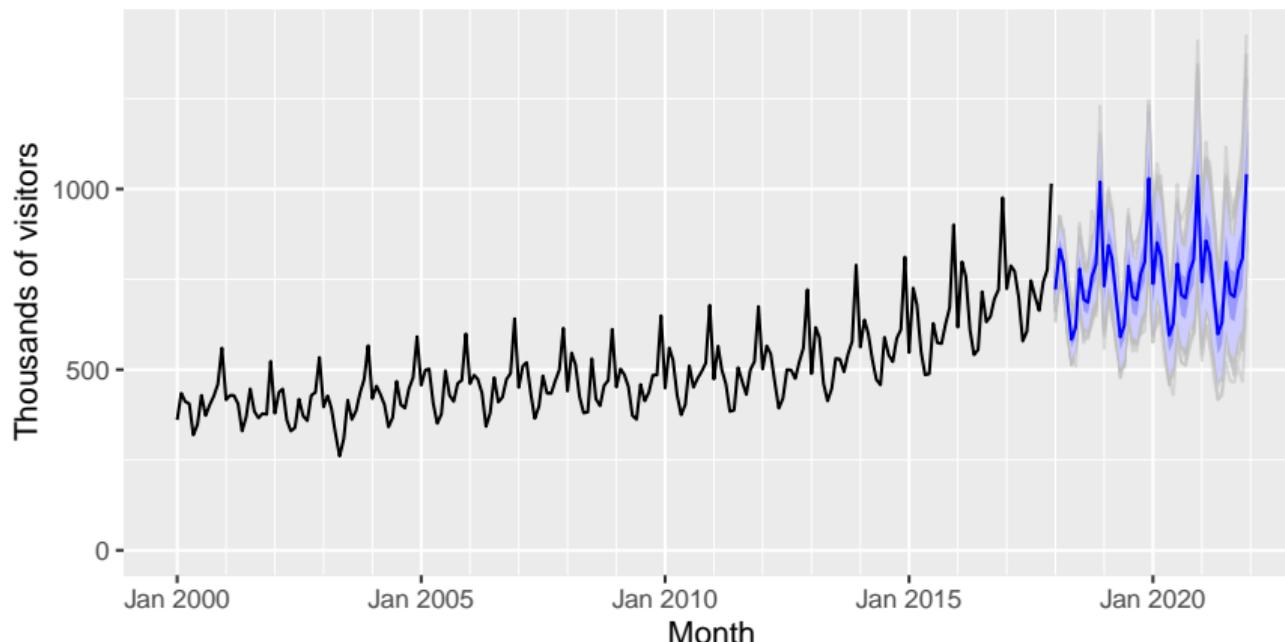


Simulated futures
from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia

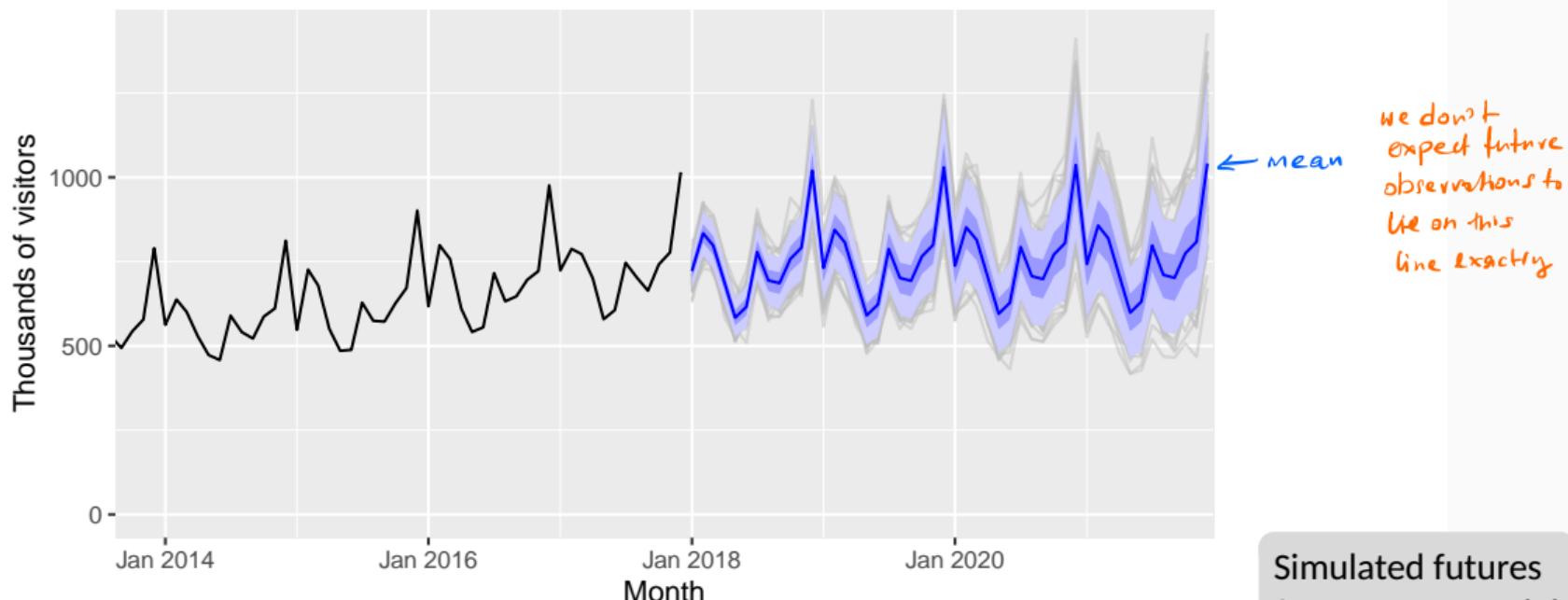


Simulated futures
from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

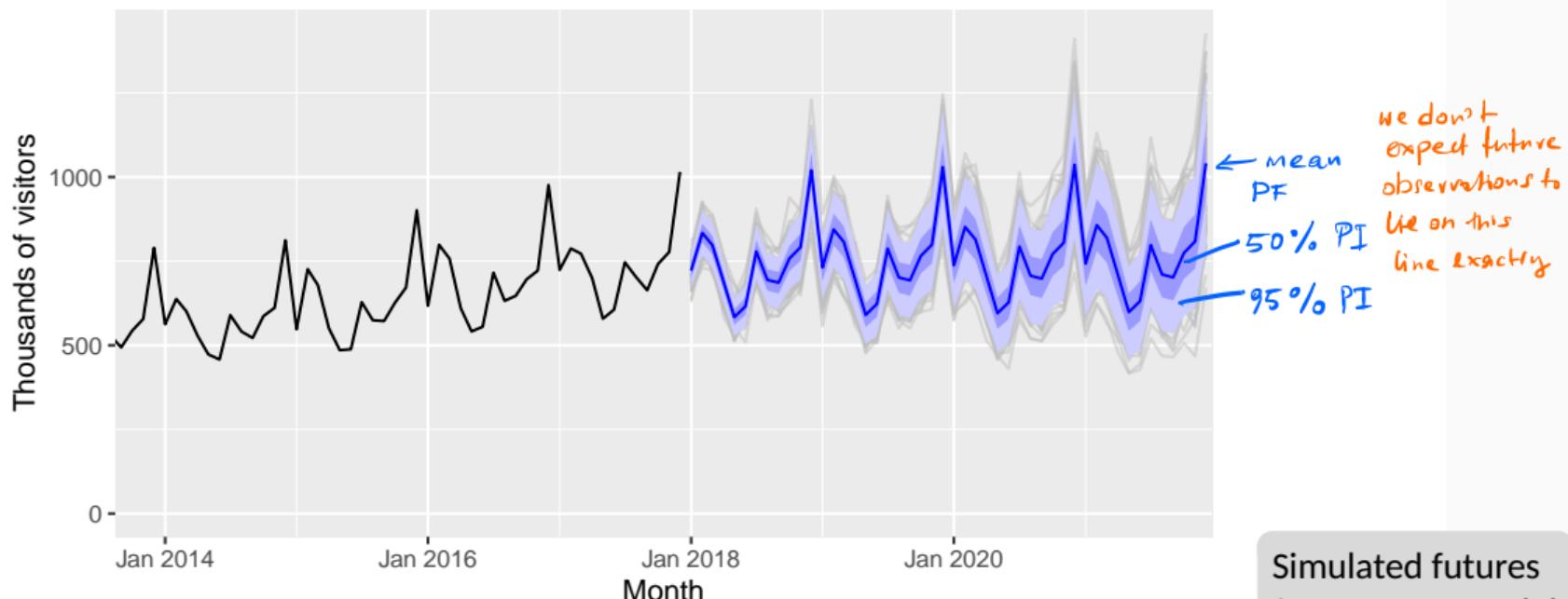
Total short-term visitors to Australia



Random futures

A forecast is an estimate of the probabilities of possible futures.

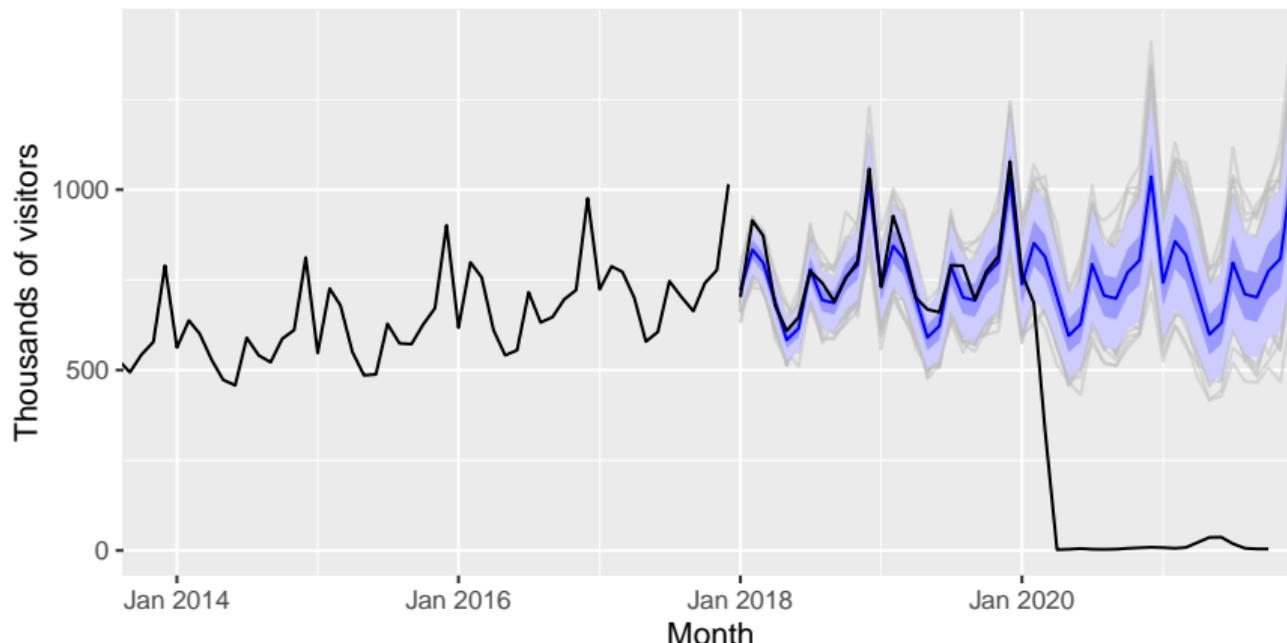
Total short-term visitors to Australia



Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia

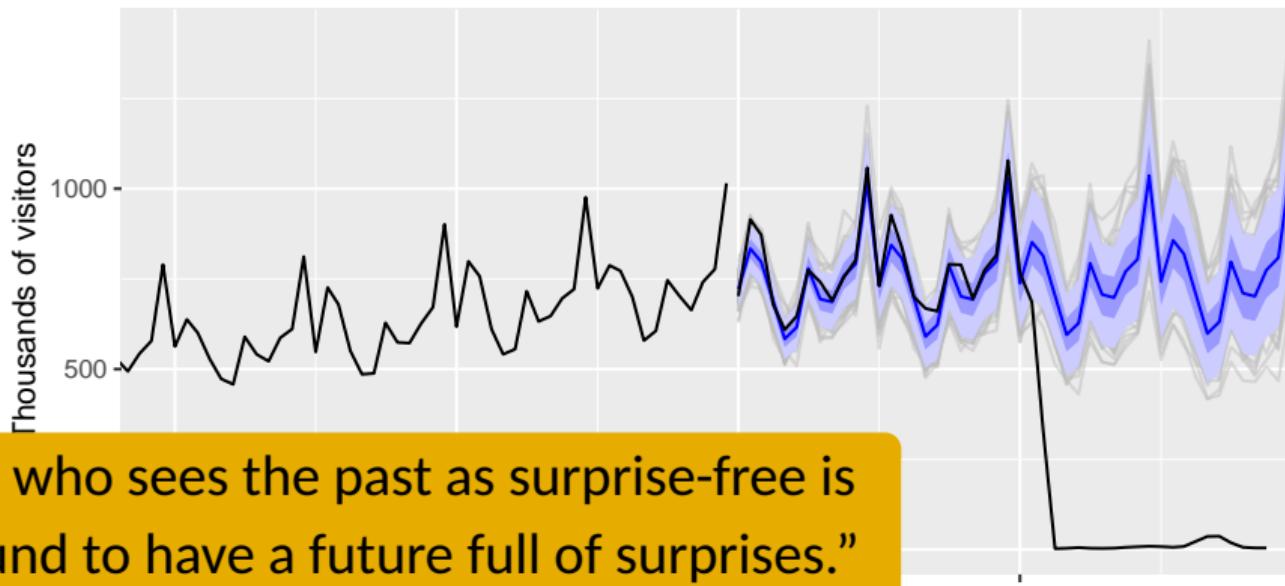


Simulated futures
from an ETS model

Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia



“He who sees the past as surprise-free is bound to have a future full of surprises.”

(Amos Tversky)

Simulated futures
from an ETS model

What is a forecast

A whole probability distribution, we call this a **forecast distribution**, which we summarise with the mean, we call this a **point forecast** and some other quantiles, we call these **prediction intervals**.

Time series in R

tsibble objects

```
global_economy
```

rows columns time index

```
## # A tsibble: 15,150 x 6 [1Y]
## # Key:   Country [263]
## #     Year Country          GDP Imports Exports Population
## #     <dbl> <fct>        <dbl>    <dbl>    <dbl>      <dbl>
## 1 1960 Afghanistan 5377777811.    7.02     4.13    8996351
## 2 1961 Afghanistan 548888896.     8.10     4.45    9166764
## 3 1962 Afghanistan 546666678.     9.35     4.88    9345868
## 4 1963 Afghanistan 751111191.    16.9     9.17    9533954
## 5 1964 Afghanistan 800000044.    18.1     8.89    9731361
## 6 1965 Afghanistan 1006666638.    21.4    11.3    9938414
## 7 1966 Afghanistan 1399999967.    18.6     8.57   10152331
## 8 1967 Afghanistan 1673333418.    14.2     6.77   10372630
## 9 1968 Afghanistan 1373333367.    15.2     8.90   10604346
## 10 1969 Afghanistan 1408888922.    15.0    10.1   10854428
```

tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]
## # Key:      Country [263]
## #       Year Country          GDP Imports Exports Population
## #   Index <fct>     <dbl>    <dbl>    <dbl>     <dbl>
## 1 1960 Afghanistan 5377777811.    7.02     4.13    8996351
## 2 1961 Afghanistan 548888896.    8.10     4.45    9166764
## 3 1962 Afghanistan 546666678.    9.35     4.88    9345868
## 4 1963 Afghanistan 751111191.   16.9     9.17    9533954
## 5 1964 Afghanistan 800000044.   18.1     8.89    9731361
## 6 1965 Afghanistan 1006666638.   21.4    11.3    9938414
## 7 1966 Afghanistan 1399999967.   18.6     8.57   10152331
## 8 1967 Afghanistan 1673333418.   14.2     6.77   10372630
## 9 1968 Afghanistan 1373333367.   15.2     8.90   10604346
## 10 1969 Afghanistan 1408888922.   15.0    10.1    10854428
```

tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]
## # Key: Year Country [263]
## #     Year Country          GDP Imports Exports Population
## #   Index  Key      <dbl>    <dbl>    <dbl>      <dbl>
## # 1 1960 Afghanistan 537777811.    7.02     4.13    8996351
## # 2 1961 Afghanistan 548888896.    8.10     4.45    9166764
## # 3 1962 Afghanistan 546666678.    9.35     4.88    9345868
## # 4 1963 Afghanistan 751111191.   16.9     9.17    9533954
## # 5 1964 Afghanistan 800000044.   18.1     8.89    9731361
## # 6 1965 Afghanistan 1006666638.   21.4    11.3    9938414
## # 7 1966 Afghanistan 1399999967.   18.6     8.57   10152331
## # 8 1967 Afghanistan 1673333418.   14.2     6.77   10372630
## # 9 1968 Afghanistan 1373333367.   15.2     8.90   10604346
## # 10 1969 Afghanistan 1408888922.   15.0    10.1   10854428
```

tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]
```

```
## # Key: Country [263]
```

```
##   Year Country      GDP Imports Exports Population
```

```
##   Index Key       Measured variables
```

```
## 1 1960 Afghanistan 537777811.    7.02   4.13  8996351
```

```
## 2 1961 Afghanistan 548888896.    8.10   4.45  9166764
```

```
## 3 1962 Afghanistan 546666678.    9.35   4.88  9345868
```

```
## 4 1963 Afghanistan 751111191.   16.9    9.17  9533954
```

```
## 5 1964 Afghanistan 800000044.   18.1    8.89  9731361
```

```
## 6 1965 Afghanistan 1006666638.  21.4   11.3   9938414
```

```
## 7 1966 Afghanistan 1399999967.  18.6   8.57  10152331
```

```
## 8 1967 Afghanistan 1673333418.  14.2   6.77  10372630
```

```
## 9 1968 Afghanistan 1373333367.  15.2   8.90  10604346
```

```
## 10 1969 Afghanistan 1408888922. 15.0   10.1   10854428
```

4 measured variables

tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]  
## # Key: Country [263]
```

A unique observation on each row for the combination of key & index.

##	Year	Country	GDP	Imports	Exports	Population
			##	Index	Key	Measured variables
## 1	1960	Afghanistan	537777811.	7.02	4.13	8996351
## 2	1961	Afghanistan	548888896.	8.10	4.45	9166764
## 3	1962	Afghanistan	546666678.	9.35	4.88	9345868
## 4	1963	Afghanistan	751111191.	16.9	9.17	9533954
## 5	1964	Afghanistan	800000044.	18.1	8.89	9731361
## 6	1965	Afghanistan	1006666638.	21.4	11.3	9938414
## 7	1966	Afghanistan	1399999967.	18.6	8.57	10152331
## 8	1967	Afghanistan	1673333418.	14.2	6.77	10372630
## 9	1968	Afghanistan	1373333367.	15.2	8.90	10604346
## 10	1969	Afghanistan	1408888922.	15.0	10.1	10854428

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]      rows   cols   index  3 key variables
## # Key:   Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##       <qtr>  <chr>    <chr>  <chr>    <dbl>
## 1 1998   Q1 Adelaide SA Business  135.
## 2 1998   Q2 Adelaide SA Business  110.
## 3 1998   Q3 Adelaide SA Business  166.
## 4 1998   Q4 Adelaide SA Business  127.
## 5 1999   Q1 Adelaide SA Business  137.
## 6 1999   Q2 Adelaide SA Business  200.
## 7 1999   Q3 Adelaide SA Business  169.
## 8 1999   Q4 Adelaide SA Business  134.
## 9 2000   Q1 Adelaide SA Business  154.
## 10 2000  Q2 Adelaide SA Business  169.
```

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##   Index     <chr>    <chr>  <chr>    <dbl>
## 1 1998   Q1 Adelaide SA  Business  135.
## 2 1998   Q2 Adelaide SA  Business  110.
## 3 1998   Q3 Adelaide SA  Business  166.
## 4 1998   Q4 Adelaide SA  Business  127.
## 5 1999   Q1 Adelaide SA  Business  137.
## 6 1999   Q2 Adelaide SA  Business  200.
## 7 1999   Q3 Adelaide SA  Business  169.
## 8 1999   Q4 Adelaide SA  Business  134.
## 9 2000   Q1 Adelaide SA  Business  154.
## 10 2000  Q2 Adelaide SA  Business  169.
```

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##   Index     Keys          <dbl>
## 1 1998 Q1 Adelaide SA Business 135.
## 2 1998 Q2 Adelaide SA Business 110.
## 3 1998 Q3 Adelaide SA Business 166.
## 4 1998 Q4 Adelaide SA Business 127.
## 5 1999 Q1 Adelaide SA Business 137.
## 6 1999 Q2 Adelaide SA Business 200.
## 7 1999 Q3 Adelaide SA Business 169.
## 8 1999 Q4 Adelaide SA Business 134.
## 9 2000 Q1 Adelaide SA Business 154.
## 10 2000 Q2 Adelaide SA Business 169.
```

Combinations of keys

76 Regions (within 8 States)

4 Purposes of Travel

$$76 \times 4 = 304$$

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##   Index    Keys          Measure
## 1 1998 Q1 Adelaide SA Business 135.
## 2 1998 Q2 Adelaide SA Business 110.
## 3 1998 Q3 Adelaide SA Business 166.
## 4 1998 Q4 Adelaide SA Business 127.
## 5 1999 Q1 Adelaide SA Business 137.
## 6 1999 Q2 Adelaide SA Business 200.
## 7 1999 Q3 Adelaide SA Business 169.
## 8 1999 Q4 Adelaide SA Business 134.
## 9 2000 Q1 Adelaide SA Business 154.
## 10 2000 Q2 Adelaide SA Business 169.
```

1 variable

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
## # Quarter Region  State Purpose Trips
## # Index     Keys          Measure
## 1 1998 Q1 Adelaide SA Business 135.
## 2 1998 Q2 Adelaide SA Business 110.
## 3 1998 Q3 Adelaide SA Business 166.
## 4 1998 Q4 Adelaide SA Business 127.
## 5 1999 Q1 Adelaide SA Business 137.
## 6 1999 Q2 Adelaide SA Business 200.
## 7 1999 Q3 Adelaide SA Business 169.
## 8 1999 Q4 Adelaide SA Business 134.
## 9 2000 Q1 Adelaide SA Business 154.
## 10 2000 Q2 Adelaide SA Business 169.
```

Domestic visitor nights in thousands by state/region and purpose of travel.

tsibble objects

- A tibble is a `data.frame` that contains a rectangular set of data.
 - ▶ Each column contains a variable (can be of different type).
 - ▶ Each row contains an observation.

tsibble objects

- A tibble is a `data.frame` that contains a rectangular set of data.
 - ▶ Each column contains a variable (can be of different type).
 - ▶ Each row contains an observation.
- A tsibble allows storage and manipulation of multiple time series in R.
 - ▶ Index: contains time information about the observation.
 - ▶ Key variable(s): optional unique identifiers for each series.
 - ▶ Measured variable(s): numbers of interest.

one column
must be an
index.
- It works with tidyverse functions.

The tsibble index

Common time index variables can be created with these functions:

Frequency	Function
Annual	<code>start:end</code>
Quarterly	<code>yearquarter()</code>
Monthly	<code>yearmonth()</code>
Weekly	<code>yearweek()</code>
Daily	<code>as_date(), ymd()</code>
Sub-daily	<code>as_datetime()</code>