

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 online: Tuesday 3-4pm (subject to changes).
- ▶ All general discussion questions will be answered on the discussion forum: <https://edstem.org/au/courses/21006/discussion> (check for answers before you ask).
- ▶ Assignment consultations - see your tutor or post to the forum.

Lecturer: Professor George Athanasopoulos

■ Contact details

- ▶ Room H5.83, Building H, Caulfield.
- ▶ Consultation online: Tuesday 3-4pm (subject to changes).
- ▶ All general discussion questions will be answered on the discussion forum: <https://edstem.org/au/courses/21006/discussion> (check for answers before you ask).
I will not consult via email.
- ▶ Assignment consultations - see your tutor or post to the forum.

Lecturer: Professor George Athanasopoulos

■ Contact details

- ▶ Room H5.83, Building H, Caulfield.
- ▶ Consultation online: Tuesday 3-4pm (subject to changes).
- ▶ All general discussion questions will be answered on the discussion forum: <https://edstem.org/au/courses/21006/discussion> (check for answers before you ask). I will not consult via email.
- ▶ Assignment consultations - see your tutor or post to the forum.

- Seminars (10:00-10:50) and Lectorials (11:00-11:50), in-person, every Tuesday, Room K321.
- Tutorials in-person.

- Joan Tan (Head Tutor)
- Ari Handayani
- Yuru (Christina) Sun
- Kulan Ranasinghe

- Joan Tan (Head Tutor)
- Ari Handayani
- Yuru (Christina) Sun
- Kulan Ranasinghe

- Consultations: available Monday to Friday (see:
<https://bf.numbat.space/>).

- Joan Tan (Head Tutor)
- Ari Handayani
- Yuru (Christina) Sun
- Kulan Ranasinghe

- Consultations: available Monday to Friday (see:
<https://bf.numbat.space/>).

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.
- Past President / Director: [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 *You will also do in your Assignment*
- Australian tourism (latest is post-Covid19) *Funded by Tourism Australia.*
- Hospital admissions (UK and Mornington Peninsula) *Through covid for Monash Peninsula.*
- Monash student enrollment numbers
- Australian prison populations *> BoCSAR*
- Macroeconomic variables *> Australian Treasury*
- Restaurant bookings *> Tatwan*
- Forecasting time series connected by aggregation constraints
(very large data) *Google, Walmart, Energy sector (AEMO), Tourism Austr., UK emergency sector, Bank of New York Mellon, etc.*

Unit objectives

*200 students in ETF3231 / 5231 + 300 in ETC3550 / 5550 at Clayton

- 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. * not a coding course
- 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-class preparation:** watch recorded videos embedded in the textbook at <http://OTexts.org/fpp3/> and read the book sections. Allow 60 minutes to do this.
- **Tuesday 10:00-10:50. In person seminar.** Review the important aspects of theory and enhance with deeper explanations or proofs when required and examples with coding. Aim: as interactive workshop as possible.
- **Tuesday 11:00-11:50. In person lectorial.** We will be going through example exercises and exam style questions. You will be practicing with me.
- Tutorials will help you with assignments. Lectorials will help with exam preparation.
→ we will do all of last year's exam.

Teaching and learning approach

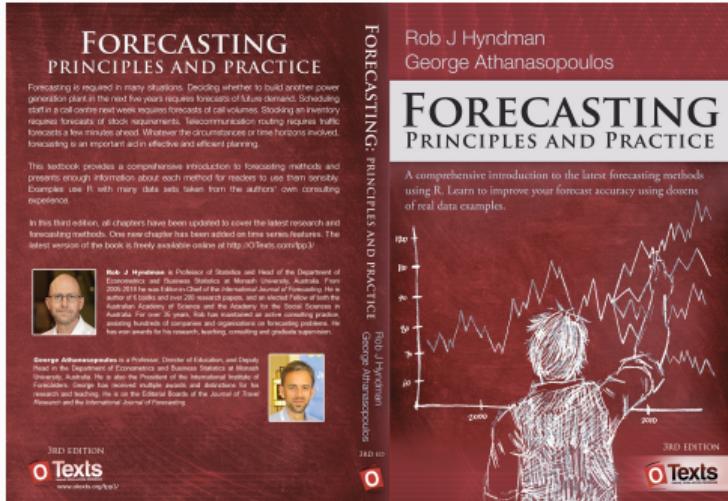
- **Pre-class preparation:** watch recorded videos embedded in the textbook at <http://OTexts.org/fpp3/> and read the book sections. Allow 60 minutes to do this.
- **Tuesday 10:00-10:50. In person seminar.** Review the important aspects of theory and enhance with deeper explanations or proofs when required and examples with coding. Aim: as interactive workshop as possible.
- **Tuesday 11:00-11:50. In person lectorial.** We will be going through example exercises and exam style questions. You will be practicing with me.
- Tutorials will help you with assignments. Lectorials will help with exam preparation.

Have R installed in your laptops and ready to go. Get help in this week's tutorials if you need to. Update R, RStudio and packages.

Textbook - key reference

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

- <http://OTexts.org/fpp3/>
- Free online
 - show translations
 - Edition three
 - Chinese.
- Printed version available [here](#)
- Data sets in associated package.
- R code for examples
 - + Pre-recorded videos → top of each section (all we will cover in the course)



Software



Available for download from CRAN: <https://cran.r-project.org>



Available for download from RStudio:
<https://posit.co/download/rstudio-desktop/>

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://posit.co/download/rstudio-desktop/>

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)
```

• We'll demonstrate this
in the second hour.

Week 1 homework

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

Week 1 homework

- Install/update R, RStudio and required packages
- See <https://otexts.com/fpp3/appendix-using-r.html>
- `install.packages(c("tidyverse", "fpp3", "GGally"), dependencies = TRUE)`
- Work through **Getting started** (5 modules) and **Writing Documents** of StartR at <https://startr.numbat.space/>
- Read Chapter 1 of the textbook and watch all embedded videos.
Pay particular attention to [Section 1.7](#).
- Read [Section 2.1](#) of the textbook and watch the embedded video.

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

- ETF3231+ETF5231: 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.
 - IA1 already posted. Will announce shortly.

Assignment schedule

Cohort	Week	Assessment task	Weight	Expectation: full marks
ETF3231+ETF5231	2	IA1	5%	Aim: get you started
ETF5231	4	GA1	5%	
ETF3231+ETF5231	6	IA2	7%	
ETF5231	7	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 or 28%) and from group assignments (weight 0.3 or 12%).
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, past exams, etc.
- Ed discussion forum for asking questions, getting help from teaching team and the bot, etc.
 - trained on the book (so better answers than ChatGPT, Cloud, etc.)
- Assignment submissions through moodle (links in the bf webpage).
- A common question: are the lectures recorded? Yes but...

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

- Includes all lecture note handouts, R code, assignments, past exams, etc.
- [Ed discussion forum](#) for asking questions, getting help from teaching team and the bot, etc.
- Assignment submissions through moodle (links in the bf webpage).
- A common question: are the lectures recorded? Yes but...

Let's explore the website.

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>

Go to <https://bf.numbat.space/assignments/A1.html>.

- Comments :
1. Forecasting competition
 2. should not take you longer than one or two.
 3. I expect everyone to get full marks.

IA1: 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:

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

$y = 90$

$$100 - 80 = 20 \quad (80 - 90)_+ = 0 \quad (90 - 100)_+ = 0 \quad IS = 20$$
$$120 - 100 = 20 \quad (100 - 90)_+ = 10 \quad (90 - 120)_+ = 0 \quad IS = 120$$

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

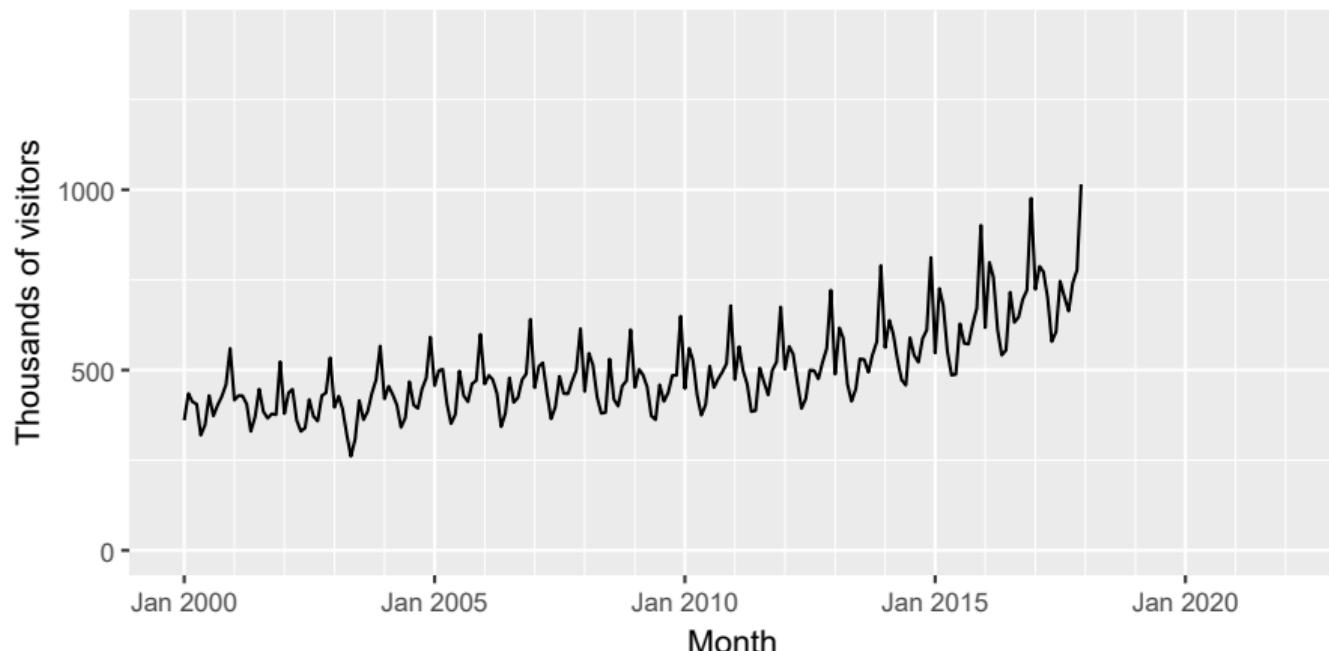
What is a forecast?

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

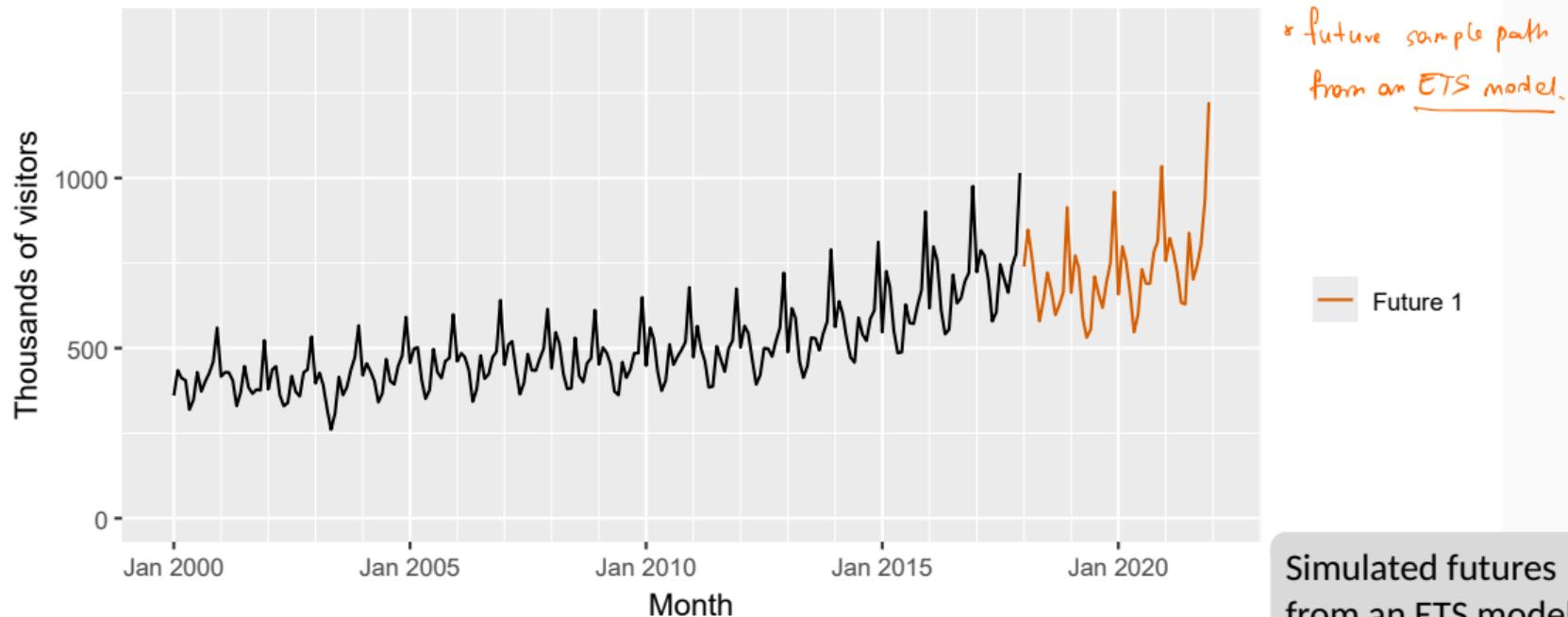


- * The future is unknown
- < We think about it as a random variable
- △ We are trying to estimate probability of a random variable.

Random futures

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

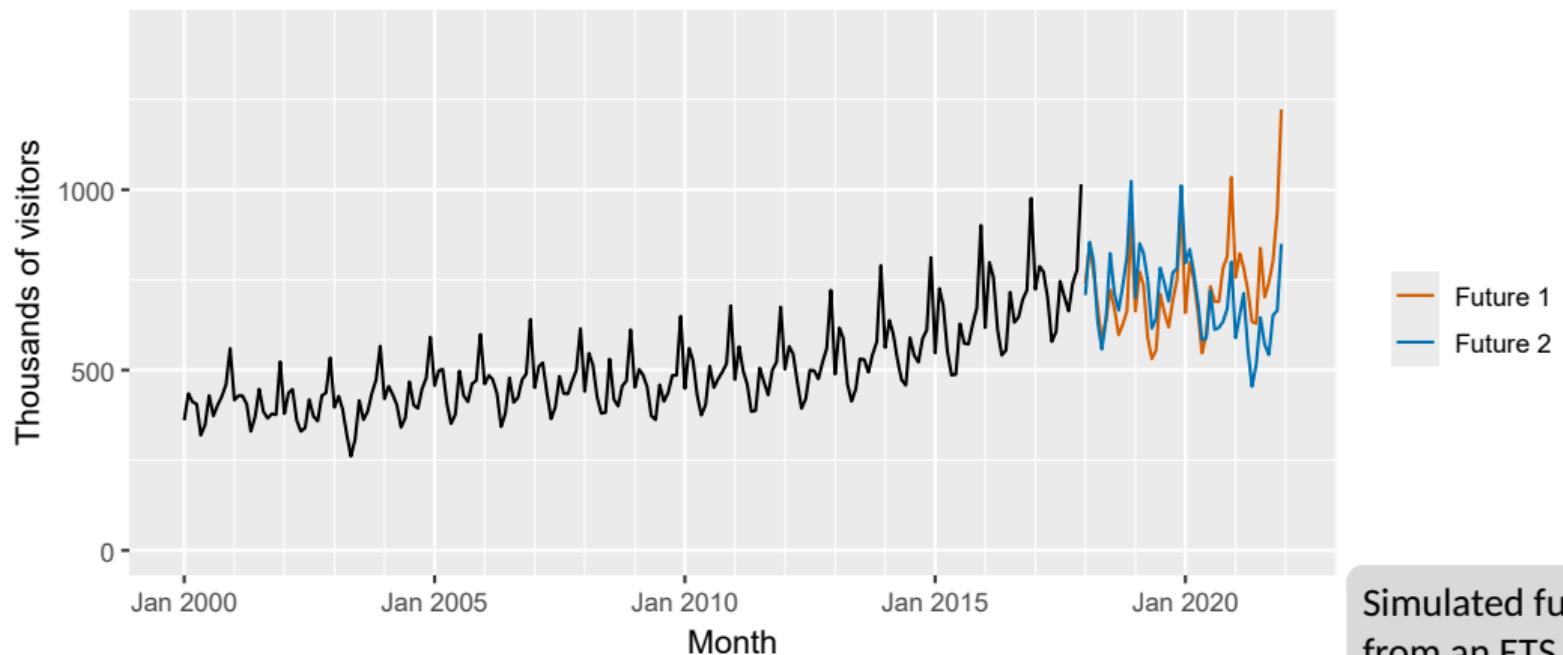
Total short-term visitors to Australia



Random futures

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

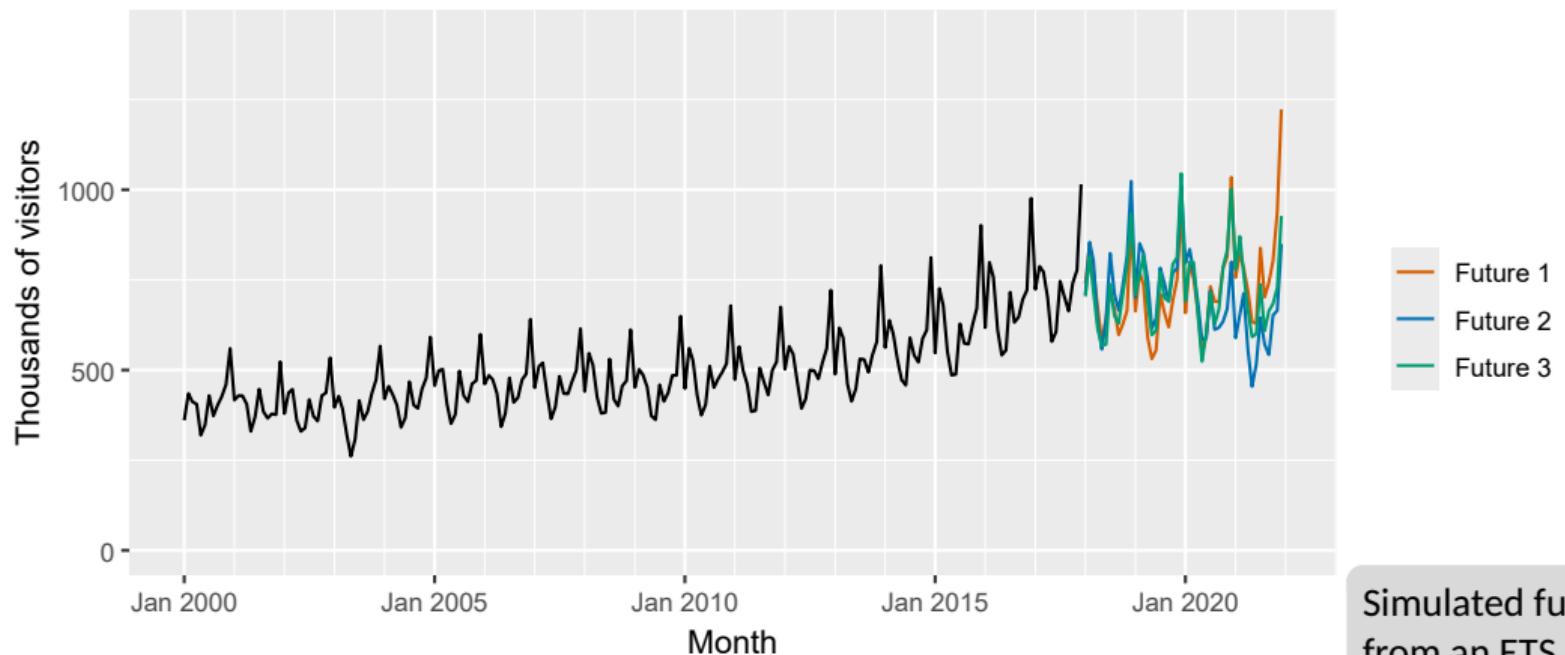
Total short-term visitors to Australia



Random futures

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

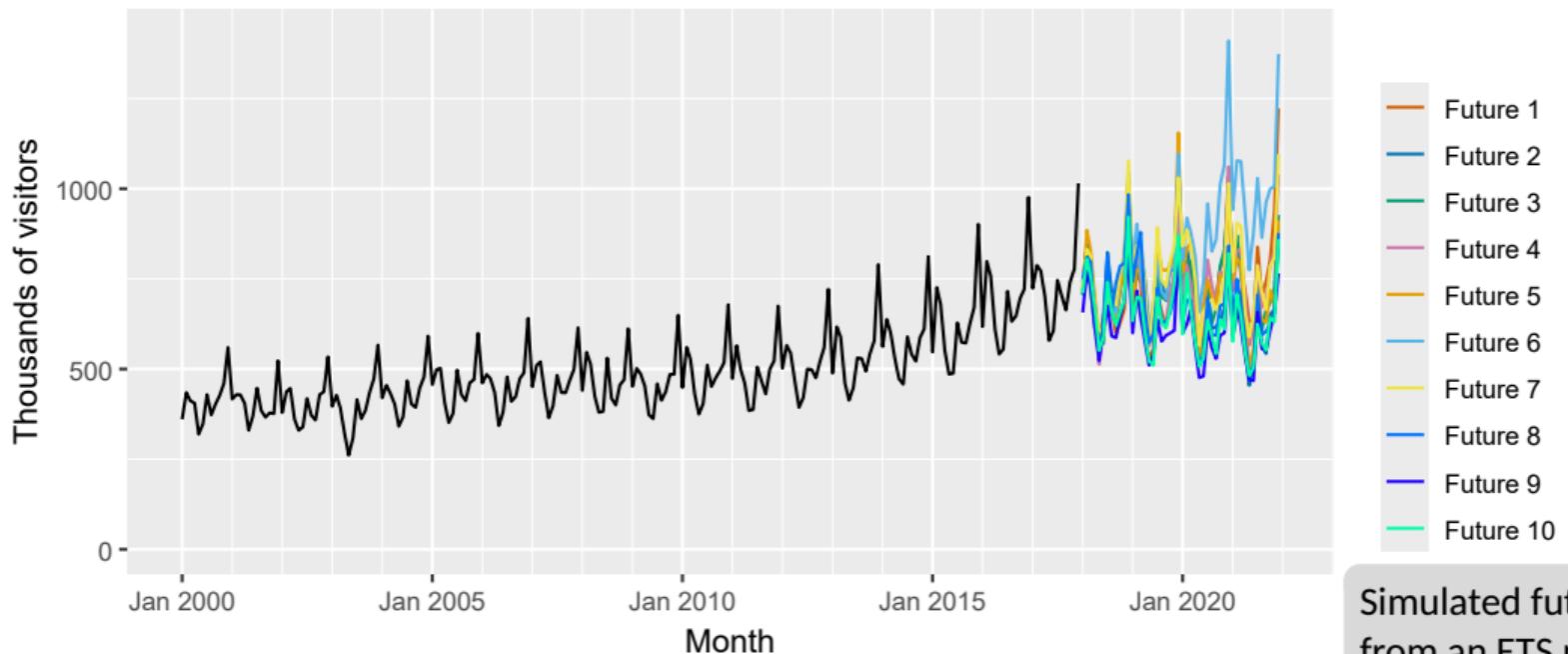
Total short-term visitors to Australia



Random futures

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

Total short-term visitors to Australia

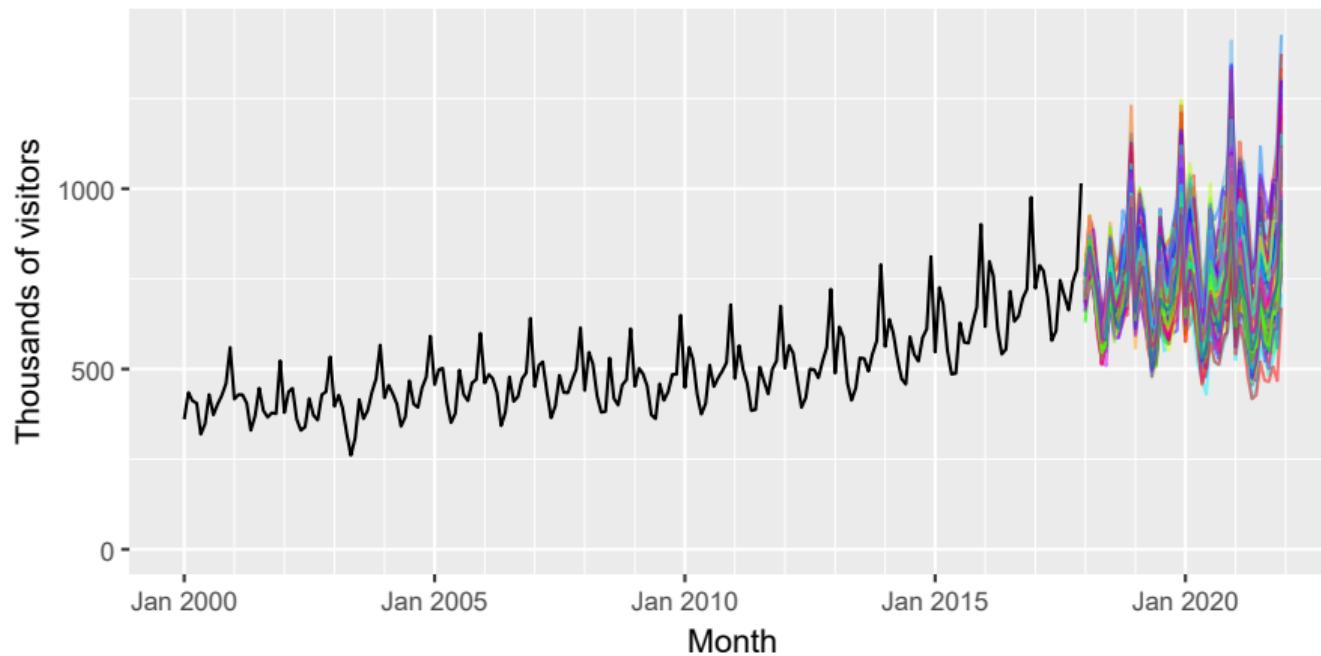


Simulated futures
from an ETS model

Random futures

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

Total short-term visitors to Australia

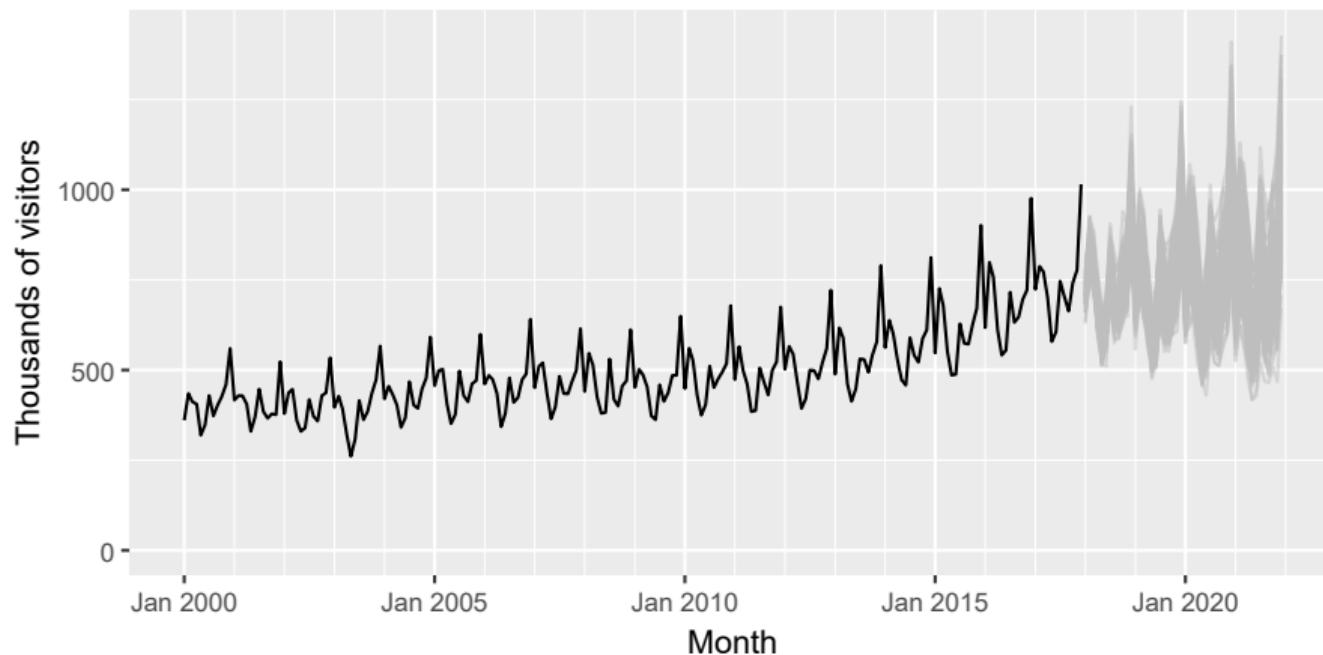


Simulated futures
from an ETS model

Random futures

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

Total short-term visitors to Australia

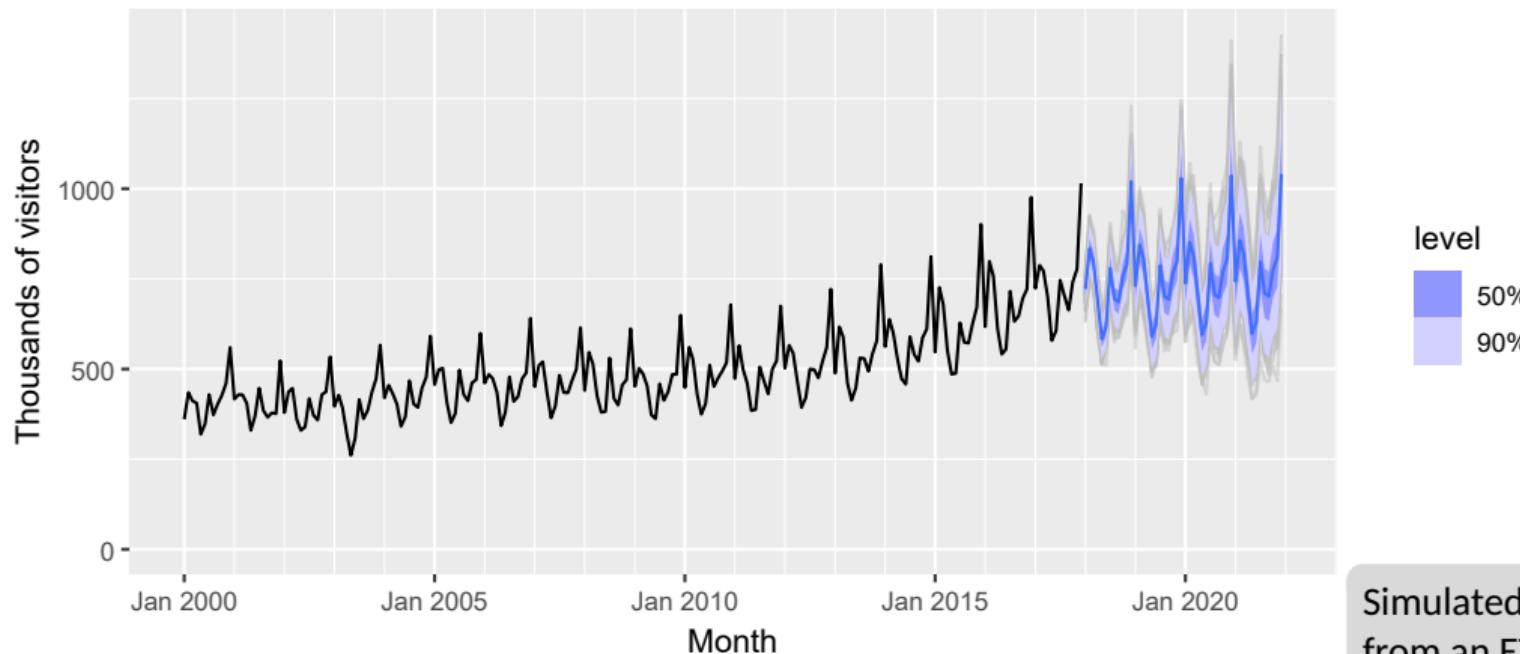


Simulated futures
from an ETS model

Random futures

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

Total short-term visitors to Australia

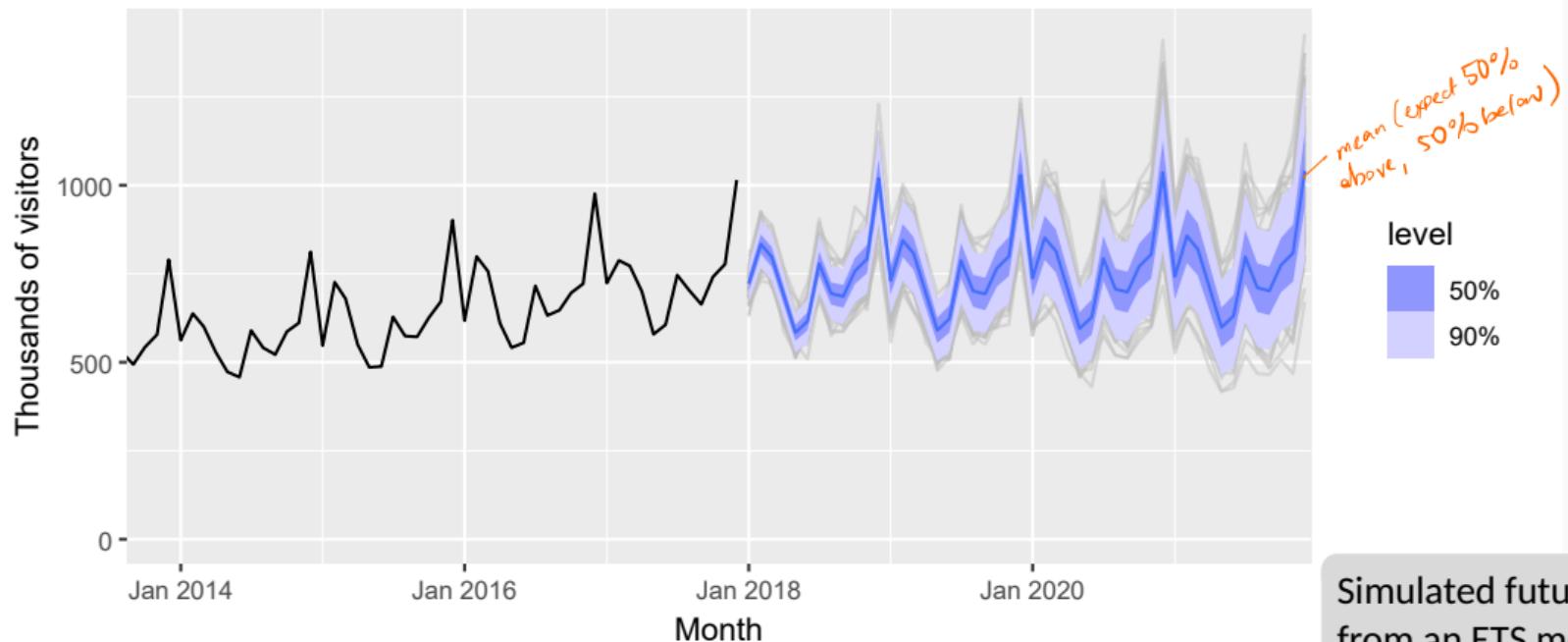


Simulated futures
from an ETS model

Random futures

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

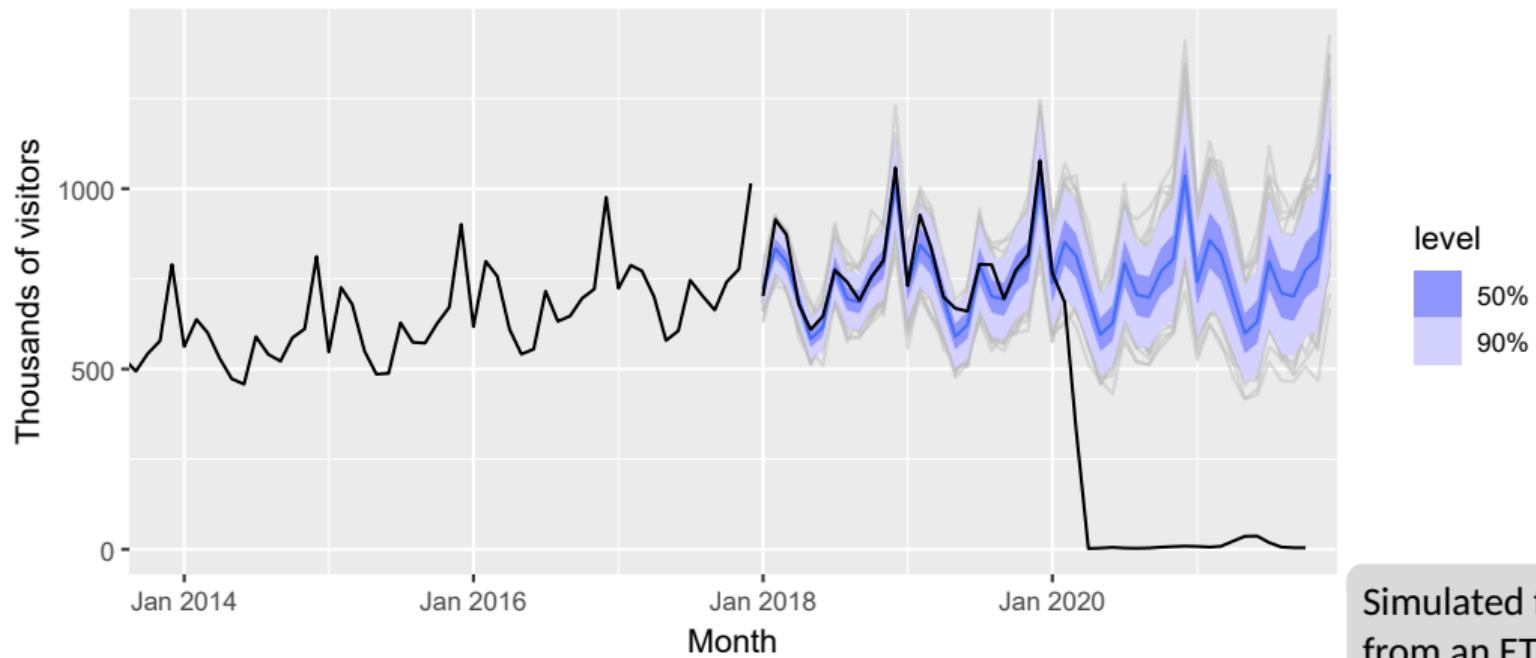
Total short-term visitors to Australia



Random futures

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

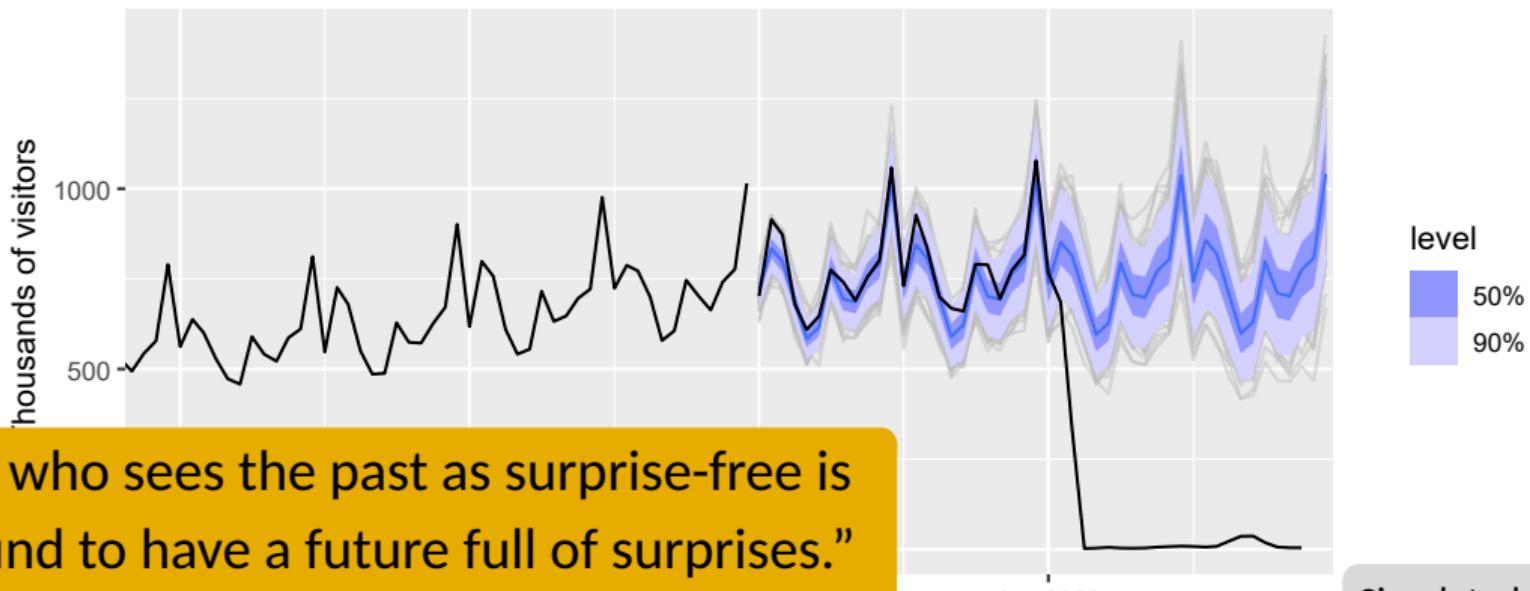
Total short-term visitors to Australia



Random futures

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

Total short-term visitors to Australia



“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**.

tsibble objects

TIME SERIES IN R

global_economy

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

rows columns
frequency
categories in the key variable(s)

	Year	Country	GDP	Imports	Exports	Population
	<dbl>	<fct>	<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	<fct>		<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	<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

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]
```

combinations of key variables

```
# Key: Region, State, Purpose [304]
```

	Quarter	Region	State	Purpose	Trips
	<qtr>	<chr>	<chr>	<chr>	<dbl>
1	1998 Q1	Adelaide	South Australia	Business	135.
2	1998 Q2	Adelaide	South Australia	Business	110.
3	1998 Q3	Adelaide	South Australia	Business	166.
4	1998 Q4	Adelaide	South Australia	Business	127.
5	1999 Q1	Adelaide	South Australia	Business	137.
6	1999 Q2	Adelaide	South Australia	Business	200.
7	1999 Q3	Adelaide	South Australia	Business	169.
8	1999 Q4	Adelaide	South Australia	Business	134.
9	2000 Q1	Adelaide	South Australia	Business	154.
10	2000 Q2	Adelaide	South Australia	Business	169.

each row is unique
across index & keys

tsibble objects

tourism

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

tsibble objects

tourism

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

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	South Australia	Business	135.
2	1998 Q2	Adelaide	South Australia	Business	110.
3	1998 Q3	Adelaide	South Australia	Business	166.
4	1998 Q4	Adelaide	South Australia	Business	127.
5	1999 Q1	Adelaide	South Australia	Business	137.
6	1999 Q2	Adelaide	South Australia	Business	200.
7	1999 Q3	Adelaide	South Australia	Business	169.
8	1999 Q4	Adelaide	South Australia	Business	134.
9	2000 Q1	Adelaide	South Australia	Business	154.
10	2000 Q2	Adelaide	South Australia	Business	169.

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	South Australia	Business	135.
2	1998 Q2	Adelaide	South Australia	Business	110.
3	1998 Q3	Adelaide	South Australia	Business	166.
4	1998 Q4	Adelaide	South Australia	Business	127.
5	1999 Q1	Adelaide	South Australia	Business	137.
6	1999 Q2	Adelaide	South Australia	Business	200.
7	1999 Q3	Adelaide	South Australia	Business	169.
8	1999 Q4	Adelaide	South Australia	Business	134.
9	2000 Q1	Adelaide	South Australia	Business	154.
10	2000 Q2	Adelaide	South Australia	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.
- 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>

very useful for
reading data
from excel or
csv files

Do example on
pmson data.