Complete the code to read in the tourism dataset from data/tourism.csv and convert it to a tsibble. It should look like tsibble::tourism when you're done.

{r} library(readr) library(tsibble) library(dplyr) library(lubridate) tourism <- read\_csv("data/tourism.csv") tourism

{r} tourism <- tourism |> mutate(Quarter = yearquarter(Quarter)) |> as\_tsibble(index = Quarter, key = c(Region,State,Purpose)) tourism

Create an **annual** time series of the Purpose of travel for visitors to Australia (summing over State and Region).

{r} tourism |> group\_by(Region, State)|> index\_by(Year = year(Quarter))|> summarise(Trips = sum(Trips))

Which Purpose of travel is most common in each state? What can you conclude about tourism in Australia?

{r} tourism |> as\_tibble() |> group\_by(Region, State, Purpose) |> summarise(Trips = sum(Trips)) |> # The code below will find the Purpose matching the most trips for each State group\_by(State) |> summarise(Purpose[which.max(Trips)])

Produce several time plots of the data. Which patterns can you observe?

{r} library(fpp3) tourism |> filter(Region == "Melbourne") |> autoplot(Trips) tourism |> filter(Region == "Melbourne", Purpose == "Business") |> autoplot(Trips) tourism |> group\_by(Purpose) |> summarise(Trips = sum(Trips)) |> autoplot(Trips) tourism |> filter(State == "ACT") |> group\_by(State) |> summarise(Trips = sum(Trips)) |> autoplot(Trips)

Create a seasonal plot for the total holiday travel to Australia over time. In which quarter is holiday travel highest and lowest?

{r} tourism|> distinct(Purpose)

{r} tourism |> filter(Purpose == "Visiting") |> summarise(Trips = sum(Trips)) |> gg\_season(Trips)

Create a seasonal subseries plot for the total business travel to Victoria over time. Does the seasonal pattern change over time?

{r} tourism |> filter(Purpose == "Business") |> summarise(Trips = sum(Trips)) |> gg\_subseries(Trips)

**Apply**

**Import vaccine\_adminstrated.csv data into R**

Let's first import data and observe it. This dataset is located indata directory in the working directory:

{r} library(tidyverse) library(tsibble) library(lubridate) vaccine\_administrated <- read\_csv("data/vaccine\_adminstrated.csv") vaccine\_administrated

After importing your data, it's important to double-check the structure of the data and the data types of the variables. The structure of the data you import into R must be a tibble/data frame.

Verify the data type of each variable in vaccine\_administered. We may need to modify the data type of month, and strike. Could you explain why?

{r} vaccine\_administrated <- vaccine\_administrated |> mutate(month = yearmonth(month), strike = as\_factor(strike)) vaccine\_administrated

**Prepare your data**

Preparing your temporal data for time series analysis and forecasting may involve several steps, such as addressing data quality issues, handling missing values, fixing duplications, creating a tsibble, and checking/filling temporal gaps.

**Check missing values and fix them**

It's essential to always check for any missing values and address them before creating a tsibble. The easiest way to know if there is a missing values is to use anyNA().

{r} anyNA(vaccine\_administrated)

The dataset you use may contain missing values (NA in R) and other data quality issues. While checking and fixing data quality issues are crucial for any project, addressing such issues goes beyond the scope of this training. For further information, you may read more about data quality issues [here](https://bookdown.org/martin_shepperd/ModernDataBook/Chap5DataCleaning.html)

**Check duplications and fix it**

Before you move forward, it's important to consistently examine your data for any duplicated observations. If such duplications are identified, address them before moving forward.

Write R code to check for duplicate observations in the data.

{r} library(janitor) get\_dupes(vaccine\_administrated)

Have you identified any duplicate observations? If your data contains no duplicates, you may proceed with creating a tsibble. Otherwise, you need to fix it.

How do you fix duplicated observations?

**Create tsibble**

You can store your data in a tsibble format, which is suitable for time series analysis and forecasting. Most functions that you use for time series analysis and forecasting, requires your data to be in a tsibble format.

Complete the following R chunk to create a tsibble.

{r} vaccine\_administrated\_tsb <- vaccine\_administrated |> mutate(month, yearmonth(month))|> as\_tsibble(index = month, key = region) vaccine\_administrated\_tsb

**Check for temporal gaps in time and fill gaps**

After creating a tsibble, it's important to check for any temporal gaps. If gaps are found, you could scan and count them. In this case, it's crucial to fill them before proceeding. However, if you're fortunate enough not to have any temporal gaps, no action is needed.

Complete the following R chunk to check if there is any temporal gap in the data:

{r} ???(vaccine\_administrated\_tsb)#check gaps ##???(vaccine\_administrated\_tsb)# show me gaps ##???(vaccine\_administrated\_tsb)# count gaps

**Are there any temporal gaps in the data?**

At times, you may come across a dataset with temporal gaps, signifying intervals where no records exist in the temporal data. Prior to do any analysis, it's important to identify and fill those gaps. You can use the below R code if such temporal gaps are present in the dataset.

{r} #If there is any gap, then fill it using fill\_gaps. #vaccine\_administrated\_tsb <- vaccine\_administrated\_tsb |> fill\_gaps(???)

**Manipulating time series**

Depending on the decision your forecast will inform, you may need to manipulate your tsibble. For instance, if you're forecasting the total dose administered in the country, or if you need to forecast quarterly doses administered for each region or for the entire country, you need to manipulate your time series frist.

Remember, you can use index\_by(), group\_by() or group\_by\_key(), and summarise() to create different temporal granularity.

Complete the following code to create total dose administrated in the country.

{r} vaccine\_administrated\_total <- vaccine\_administrated\_tsb |> group\_by(region) |> summarise(dose\_adminstrated = sum(dose\_adminstrated)) vaccine\_administrated\_total

Depending on the forecasting task on hand, you may need to work with other time granularities such as quarterly time series.

Complete the following code to create total dose administrated in the country.

{r} #| label: quarterly library(lubridate) vaccine\_administrated\_tsb |> index\_by(Quarter = yearquarter(month)) |> summarise(dose\_adminstrated = sum(dose\_adminstrated))

What if you need to create quarterly dose administrated in each region, write the R to achieve that.

{r} vaccine\_administrated\_tsb |> as\_tibble() |> group\_by(region) |> summarise(dose\_adminstrated = sum(dose\_adminstrated)) |> arrange(desc(dose\_adminstrated))

**Visualising time series**

**Time plots**

To understand your data, you can start by producing time plot of dose administrated in 9 regions.

{r} vaccine\_administrated\_tsb |> autoplot(dose\_adminstrated)

You can also focus on any region by filtering the region:

{r} vaccine\_administrated\_tsb |> filter(region == "A") |> autoplot(dose\_adminstrated)

Do you observe any systematic pattern in time series plots?

**Seasonal plots**

In time series analysis, we are looking for consistent pattern. Complete the following code to create seasonal plot to see if there is any obvious monthly consistent pattern.

{r} vaccine\_administrated\_tsb |> gg\_season(dose\_adminstrated)

It might not be easy to see the systematic pattern when you plot many time series together,instead you can first filter the time series of interest and then plot it:

{r} vaccine\_administrated\_tsb |> filter(region == "A") |> gg\_season(dose\_adminstrated)

Do you observe any consistent pattern? How different the pattern is across region?

**Seasonal subseries plot**

You might be also interested in observing how doses administrated within each month/quarter change over time, as well as understanding the dose administrated changes across different season. This could be plotted for each region separately . Complete the following code to create the plot for your region of interest:

{r} vaccine\_administrated\_tsb |> filter(region == "A") |> gg\_subseries(dose\_adminstrated)

Do you see any pattern that has not been obvious with time plot and seasonal plot?

**ACF plots**

In forecasting, we would be interested in understanding how similar a time series is to the lags of itself. We often measure this similarity by calculating the correlation (i.e. the linear association) between a time series and its lags and then plot it.

{r} vaccine\_administrated\_tsb |> ACF(dose\_adminstrated, lag\_max = 12) |> autoplot()

You can also focus on any time series of interest by filtering the region:

{r} vaccine\_administrated\_tsb |> filter(region == "G") |> ACF(dose\_adminstrated, lag\_max = 12) |> autoplot()

What can you say about the correlation between dose administrated with its lags?