

# ST5209/X Assignment 1

Due 3 Feb, 11.59pm

## Set up

1. Make sure you have the following installed on your system:  $\text{\LaTeX}$ , R4.2.2+, RStudio 2023.12+, and Quarto 1.3.450+.
2. Clone the course [repo](#).
3. Create a separate folder in the root directory of the repo, label it with your name, e.g. `yanshuo-assignments`
4. Copy the `assignment1.qmd` file over to this directory.
5. Modify the duplicated document with your solutions, writing all R code as code chunks.
6. When running code, make sure your working directory is set to be the folder with your `assignment.qmd` file, e.g. `yanshuo-assignments`. This is to ensure that all file paths are valid.<sup>1</sup>

## Submission

1. Render the document to get a `.pdf` printout.
2. Submit both the `.qmd` and `.pdf` files to Canvas.

## Question 1 (Quarto)

Read the [guide](#) on using Quarto with R and answer the following questions:

- a) Write a code chunk that imports `tidyverse` and `fpp3`.
- b) Modify the chunk so that only the following output is shown (i.e. the usual output about attaching packages and conflicts is not shown.)

```
library(tidyverse)
library(fpp3)
```

---

<sup>1</sup>You may view and set the working directory using `getwd()` and `setwd()`.

- c) Modify the chunk so that it is executed but no code is shown at all when rendered to a pdf.
- d) Modify the document so that your name is printed on it beneath the title.

### Question 2 (Livestock)

Consider the `aus_livestock` dataset loaded in the `fpp3` package.

- a) Use `filter()` to extract a time series comprising the monthly total number of pigs slaughtered in Victoria, Australia, from Jul 1972 to Dec 2018.
- b) Make a time plot of the resulting time series.

### Question 3 (Data cleaning)

Inspect the function `process_sgcpi()` located in `_code/clean_data.R`. This function is used to convert the raw Consumer Price Index (CPI) data in `_data/raw/sg-cpi.csv` into a tsibble, stored in `_data/cleaned/sgcpi.rds`.

- a) In line 9, what does `skip = 10` and `n_max = 152` do? Why do we need to do this when reading the csv file?
- b) In line 14, what does `t()` do? Why do we need to do this in order to make a tsibble?

### Question 4 (Beer production)

Consider the `aus_production` dataset loaded in the `fpp3` package. We will study the column measuring the production of beer.

- a) Make a time plot of the beer production time series.
- b) Describe the observed trend.
- c) Make a seasonal plot.
- d) What is the period of the seasonality?
- e) Describe the seasonal behavior.

### Question 5 (Pelts)

Consider the `pelt` dataset loaded in the `fpp3` package, which measures the Hudson Bay Company trading records for Snowshoe Hare and Canadian Lynx furs from 1845 to 1935.

- a) Plot both time series on the same axes. *Hint: Use `pivot_longer()` to create a key column.*

- b) What happens when you try to use `gg_season()` to the lynx fur time series? What is producing the error?
- c) Make a lag plot with the first 20 lags. Which lags display strong positive correlation? Which lags display strong negative correlation? Verify this with the time plot.
- d) If you were to guess the seasonality period based on the lag plot, what would it be?
- e) Use the provided function `'gg_custom_season()' in _code/plot_util.R`<sup>2</sup> to make a seasonal plot for lynx furs with the period that you guessed.<sup>3</sup> Does the resulting plot suggest seasonality? Why or why not?

### Question 6 (Box-Cox, Q3.3 in FPP)

Why is the Box-Cox transform unhelpful for the `canadian_gas` data?

---

<sup>2</sup>You can load this function using `source("../_code/plot_util.R")`.

<sup>3</sup>Unfortunately, it seems `'gg_season()' does not allow this functionality.`