# Executive Summary

‘New Zealand’s Interactive Emissions Tracker’ is a web application which is currently hosted on the Ministry for the Environment’s public website and provides a visual representation of New Zealand’s Greenhouse Gas inventory across a series of years. In order to allow better control over the delivery of information and the implementation of additional features, the Ministry is interested in transitioning from outsourcing the app into developing and maintaining it all within the Ministry.

To allow for an easier development process, the tracker was completed piece-by-piece, where new components were implemented following a logical sequence. While still lacking in some functionality to act as a complete replacement for the current Emissions Tracker, this replication succeeds as a proof of concept. It demonstrates the feasibility of developing the tracker in-house entirely using R Shiny, as well as providing a working foundation to continue development from.

Before transitioning towards publishing the in-house developed Emissions Tracker, it is recommended to make the necessary improvements to the code and implement all missing (but required) features. Furthermore, if possible, minor adjustments to the data source (removing unnecessary columns and splitting ‘time series’ to multiple different sheets’) could improve both the development process and running efficiency of the program.

Contents

[Executive Summary 1](#_Toc63259002)

[Introduction 3](#_Toc63259003)

[Set-up 3](#_Toc63259004)

[Data Cleaning 3](#_Toc63259005)

[Scoping Functionality 3](#_Toc63259006)

[Development Process 4](#_Toc63259007)

[Setting up User Interface (UI) 4](#_Toc63259008)

[Producing Required Graphs 4](#_Toc63259009)

[Reactivity 5](#_Toc63259010)

[Conclusions 5](#_Toc63259011)

[Recommendations (Steps moving forward) 6](#_Toc63259012)

[Appendix 7](#_Toc63259013)

[Appendix A 7](#_Toc63259014)

[Appendix B 8](#_Toc63259015)

[Appendix C 9](#_Toc63259016)

# Introduction

The Ministry for the Environment’s Emissions Tracker is an interactive web application which offers, using graphs and tables, a visual representation of New Zealand’s Greenhouse Gas (GHG) inventory. Being hosted on a public domain, this allows anyone to readily track New Zealand’s progress towards our climate goals and quantitatively measure the differences we’ve made along the way.

Currently, the Emissions Tracker is being outsourced to Datacom, where the Ministry provides the required data and Datacom returns with the completed product. There is a demand for this process to be completed in-house, allowing the Ministry to take complete control over the delivery of this product and more easily implement additional features at the request of Stakeholders. As such, it was tasked to recreate the currently live Emissions Tracker using R Shiny in the short term, and to continuously update and improve the tracker with any requested features in the long term.

# Setup

Due to the relatively large scope of the project, a few preliminary steps were taken to ensure a smoother development process.

## Data Cleaning

Provided was a single Excel spreadsheet which contained a detailed breakdown of the GHG emissions for each Sector from years 1990 to 2018. Several issues were noted which would have an impact on the completion of the product, including:

* Yearly breakdowns of emissions are stored on different sheets.
* First and last sheet aren’t yearly breakdowns, and thus shouldn’t be included in the same list when importing.
* Several rows in each year are specified as “Exclude” or have 0 as the value.
* Several columns in each year are unnecessary (either “Exclude” or provides redundant information).
* Time series data has a dropdown box to filter for specific GHG. R import function unable to change this, thus cannot access GHG specific time series data.

A function was created in order to import all the information while taking into consideration the issues mentioned above, producing a set of data frames (one for yearly emission breakdown data and the other for time series data) which are easy to use and fit for purpose.

An assumption was made in the creation of this function, in that every single yearly emission breakdown sheet follows the exact same structure. Columns that are to be kept are ‘hard-coded’ into the function, and any deviation can result in the final data frame to contain the incorrect information. This can have the fatal consequence of crashing the app entirely. There appears to be no problem with the current data, but this is an important assumption to keep in mind and work around for future data sets.

## Scoping Functionality

As the task is to recreate an already existing application, it was essential to have a firm understanding of the functionality for the current Emissions Tracker. This process involved exploring through every aspect of the current Emissions Tracker and noting what it is capable of.

The order of priority for the tracker’s functionality implementation was based on their relative importance. Features which are nice to have but not completely necessary were of a secondary priority, and only attempted once the core functionality was fully implemented. There are certain features which are of lower priority that have still yet to be implemented, providing clear room for improvement.

# Development Process

With the Emissions Tracker being recreated in R Shiny and personally having no prior experience with using Shiny, it was decided to slowly build up the app piece-by-piece, learning the different aspects of Shiny along the way.

## Setting up User Interface (UI)

Beginning with the most basic aspect of the Shiny app, the current Emissions Tracker follows quite a simple structure consisting of a ‘main’ panel and a ‘side’ panel. The main panel displays the emissions graph and other useful pieces of information, whereas the side panel contains two smaller graphs detailing the time-series and GHG breakdown of the main graph (*refer to Appendix A for an annotated screenshot of the layout, as well as a more detailed breakdown of the different components of the UI).*

Due to a lack of knowledge in CSS/html, an exact replication of the current tracker’s UI was impossible and slight adjustments were made. While no visual information was lost, this resulted in the dashboard to become more cluttered and less intuitive to navigate.

## Producing Required Graphs

As mentioned in the *Data Cleaning* section, a single data frame containing all the yearly emissions breakdown, *emissions,* was produced. However, there were issues in converting this data into a presentable floating bar chart as seen in the Emissions plot for the current tracker. A solution was devised, utilising *geom\_crossbar* from the *ggplot2* package which allows the starting and ending point of each bar to be manually assigned. These values were bound to a filtered down version of the *emissions* data frame, which contained only the necessary columns and observations, and is passed through *geom\_crossbar* to create our floating bar chart.

This process was applied whenever an update was required for the Emissions graph, and only require as input the observations we wish to produce the graph of.

Another obstacle faced was devising a method which was able to correctly aggregate together all subsections of a given sector/subsection. For reference, from *Figure 1*:

Figure : Sector structure reference

Figure 2: Sector breakdown reference

* **[5. Waste]** is the main sector which we are breaking down.
* **[5.A Solid Waste Disposal] and [5.B Biological Treatment of Solid Waste]** are the two direct subsections (in this reference) for our sector.
* **[5.A.1 Managed Waste Disposal Sites], [5.A.2 Unmanaged Waste Disposal Sites], [5.A.3 Uncategorized Waste Disposal Sites] and [5.B.1 Composting]** are the direct subsections of the above subsections. We should only return a data frame containing information corresponding to the subsection which is associated with its preceding subsection (i.e. if we chose [5.B Biological Treatment of Solid Waste], we should only return information for [5.B.1 Composting]).

It is essential for the method to operate correctly under all conditions, else the resulting output won’t be correct. Several attempts at this method were made with varying levels of success before finalising on the current solution. It involves recording each subsection we choose to enter and filtering this with the *Location of Node in hierarchy* column (*refer to Appendix B for a snapshot of this column and a more detailed breakdown of how this works).* By doing this, we are left with all the immediate subsections of the chosen item, allowing us to generate the required graphs.

A previous attempt involved filtering by the associated ‘codes’ (i.e. 5.A, 5.A.1, 5.A.1.b in the above figure are the codes associated with their respective subsection) to try and capture the immediate subsections of the chosen item. This however ultimately failed due to two main issues:

* Structure of the codes weren’t always consistent which resulted in the function to not operate properly.
* Certain sections had the exact same code but belonging to different subsections which the function couldn’t differentiate between, leading to an excess in the observations that are pulled into the data frame.

among a series of other edge-cases which required explicit code chunks to solve and thus decreased the efficiency of the program.

## Reactivity

Being an interactive web app, the Emissions Tracker is required to be able to take in user input and update the output based on the parameters of the input. For the tracker specifically, this takes the form of updating each of the graphs based on the mouse input of the user*,* whether it be hover or click.

Due to a lack of experience with R Shiny, there was much difficulty in understanding the different reactive expressions available and knowing which are most applicable for my desired outcome. Furthermore, with the sheer number of moving parts in the current tracker, each reactive chunk was considered independently as a means of simplifying the problem. However, this resulted in a level of code redundancy which further reduced the efficiency of the program (albeit not to a level that is significant).

For a more technical breakdown of how reactivity is applied within the app, please refer to *Appendix C.*

# Conclusions

The recreated Emissions Tracker succeeds as a proof of concept, demonstrating that it is entirely feasible to produce a product similar to the current tracker using R Shiny. The core functionality of the app itself is present, but many smaller but beneficial features are not yet implemented.

# Recommendations (Steps moving forward)

As previously mentioned, while the app works well enough as a proof of concept, there are still many improvements that need to be made before it can fully replace the current Emissions Tracker. These improvements include, but are not limited to:

* Checking for and removing any bugs in the current version.
* Remove the code redundancy within reactivity.
* Improve aesthetics of dashboard.

Additional features that are present in the current Emissions Tracker but not the recreation include:

* Only Emissions plot has a tabular view, Year and GHG graphs do not.
* Hover over a bar to obtain its value is only present in Emissions plot, not present in Year or GHG graphs.
* No way of exporting data for any of the three graphs outputs.
* Unable to play animation of year-to-year changes.
* Observations with an emission below a certain threshold show as 0, rather than <0.01.
* No way of switching between ‘Net Emissions’ and ‘Gross Emissions’ on the top level.
* Unable to select a subsection in the sector tracer and jump directly to the selected subsection.
* Only the tracker itself is present in the Shiny app. The surrounding text, ‘FAQs’, ‘Tutorial’, ‘Download Image’ and sharing hyperlinks do not exist.

If possible, it is also recommended for certain changes to be made in the way the data is stored. A large portion of the columns within the spreadsheet are listed entirely as ‘Exclude’, making them unnecessary for importing *(Refer to Appendix D for a snapshot of these columns)*. However, due to limitations with reading from excel files, we are unable to exclude these columns prior to the import and can only subset them out afterwards. This not only has a minor slow down on the import speed, but more importantly it requires us to assume there are no changes to the layout of the spreadsheet as mentioned in *Data Cleaning.*

Finally, another recommendation for ease of importing is to split up the *time series* sheet into multiple different sheets. Currently there is a dropdown box where we select the GHG we wish to see time series data for (with ‘All gases’ being the default). However, when importing into R, we aren’t able to change the values of the dropdown box, thus requiring us to apply various other functions to get our desired time series data. By expanding this dropdown box into multiple sheets, this allows the process of extracting our desired time series to be better optimised and streamlined.

# Appendix

## Appendix A

Figure : Breakdown of UI Components

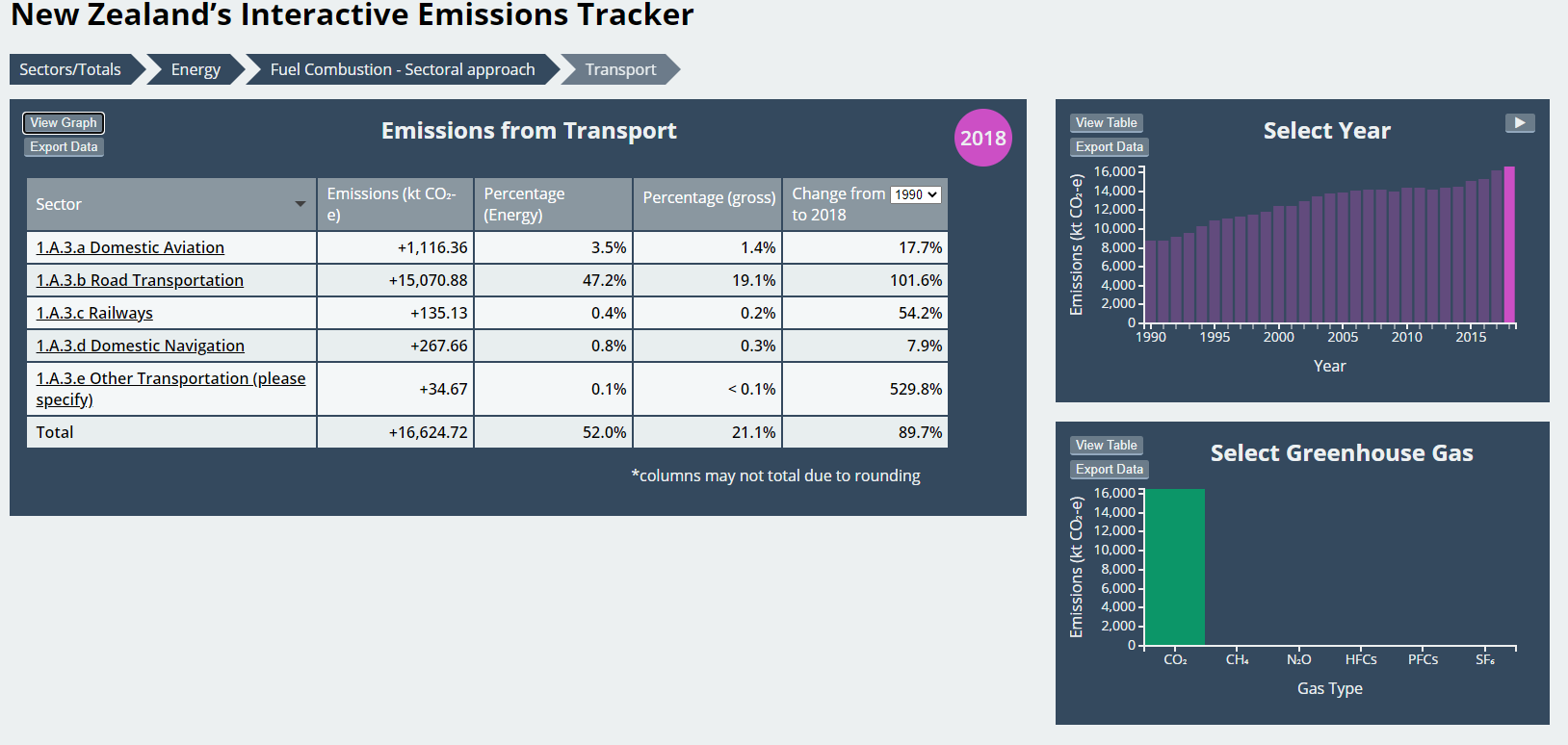
* *Main Panel* and *Side Panel* are the two columns that are used to separate the webpage.
  + *Side Panel* is further split up into 2 rows, one for each of *Year Graph* and *GHG Graph,* both of which are of *plotOutput* type.
  + *Main Panel* is further broken into smaller components, with *Emissions Graph* (of plotOuput type) taking up majority of the column.
  + *Sector Tracer* is an *uiOutput* of *selectInput* type, which tracks the order of sectors/subsections we’ve selected.
  + *Output Type* is a *radioButton*, in which we are able to swap between the graph view of the *Emissions Graph* (as shown in *Figure 3*), or a tabular view which is shown in *Figure 4.*

Figure : Tabular View of Emissions Graph

## Appendix B

Figure : Snapshot of Location of Node in Hierarchy column

As seen in *Figure 5*, *Location of Node in hierarchy* is a column which has similar information as the aforementioned *Sector Tracer* in Appendix A.   
Each observation contains the list of all the subsections that precede the current observation (i.e. to reach the node **1.A.1 Energy Industries**, we have to pass through **[Sectors/Totals], [1. Energy]** and **[1.AA Fuel Combustion – Sectoral approach],** where [Sectors/Totals] is just the indicator for the base layer). It is important to make the distinction that *Location of Node in hierarchy* does **not** include the node it is currently at, whereas *Sector Trace* does.

As such, we are able to filter *Location of Node in hierarchy* by the exact tracking information stored for *Sector Tracer*, and thus determine all of its immediate subsectors.

For example, suppose we’re currently at **[1.A.1 Energy Industries]** and need to know all immediate subsectors branching from this node to be used for plotting.   
By filtering *Location of Node in hierarchy* for the *Sector Tracer* information (which is currently **[Sectors/Totals][1. Energy][1.AA Fuel Combustion – Sectoral Approach][1.A.1 Energy Industries]**), we are left with three observations that satisfy this filter (**1.A.1.a**, **1.A.1.b** and **1.A.1.c**), each corresponding to one of the three subsectors that branch from **[1.A.1 Energy Industries]**.

## Appendix C

Two reactive expressions are used within the Shiny App to capture events, being:

* reactive
* observeEvent

These two reactive expressions, while similar, have slight differences which better gear them for different uses.

**reactive**requires an explicit call from any source to run and returns an output. However, the call to the *reactive* function is only made when the information it returns is different from what it had before. As such, it is called when the data frames storing data used for plotting changes.

**observeEvent** is called implicitly when the Shiny app registers an event from its specified event expression and does not return an output. As such, it is used mainly for changing reactive variables (i.e. *reactiveValues)* when necessary (e.g. changing the year value from the currently selected to one newly chosen from the user).

The final form of reactivity within the Shiny app comes in the form of **reactiveValues**. These are effectively global variables that trigger an event when its value changes for any functions/outputs that depend on its value. This in combination with **observeEvents** are what drives the interactive nature of the Shiny app.

When an event is registered by **observeEvent**, it results in a change in the value of a **reactiveValue.** This then triggers a separate event for the outputs, that call a **reactive** function which updates the data frames used for plotting and plots the new graphs.

## Appendix D

Figure 6: Snapshot of Columns with Exclude

Columns beyond the ‘Total GHG’ which break down the different fuel types are listed almost entirely of ‘Exclude’. Even for the few exceptions where there are values present (such as rows 17 and 25), the information is not presented anywhere in the Emissions Tracker.