

BMP Dashboard Manual

Ané Cloete

2023-10-05

Contents

1	Hi :)	5
1.1	Owners	5
2	Files within folders and folders within files and more files	7
2.1	Biodiversity KPI mapping Master	7
2.2	BMPDashboard	9
3	The Data	13
3.1	Setting up OneDrive	13
3.2	Getting the data	14
3.3	Data cleaning and preparation	16
3.4	Pulling photos and documents from OneDrive	21
3.5	Things to think about	22
4	Footnotes and citations	23
4.1	Footnotes	23
4.2	Citations	23
5	Blocks	25
5.1	Equations	25
5.2	Theorems and proofs	25
5.3	Callout blocks	25

6	Next Steps	27
6.1	Different years different data:	27
6.2	Data download and clean automation:	27
6.3	Survey Photos:	27

Chapter 1

Hi :)

Everything you need to know to takeover the maintaining and building of the BMP dashboard is in this bookdown. The first chapter will outline the R project folder structure and contents and then in Chapter two I'll talk about how to get the biodiversity data form OneDrive and prepare the data for the dashboard. The remaining chapters will delve into the code itself.

Before starting, here is a short list of handy resources:

- R
- Shiny
- Github
- Bookdown (if you want to know how to make something like this).

These are big and comprehensive resources that will aid you along the way. But remember google and stackoverflow are your best friends! And ChatGPT might become your closest and most frustrating colleague.

1.1 Owners

- Ané Cloete (2023)

If you are the new dashboard manager, add your name as well! All you need to do is open the bookdown project and navigate to the “index.Rmd” file and scroll down till you see this section. In the next chapter you'll see where this folder is located.

Remember the dashboard is work in progress, so feel free to change and modify it as you like and most importantly have fun with it! Let's begin!

Chapter 2

























Files within folders and folders within files and more files

Let's go over all the files and folders you've now gotten access to starting with the OneDrive Folder.

2.1 Biodiversity KPI mapping Master
























This is, as it's name suggests, the master folder! And all the NB things are stored here, this is what you should see (unless more folders have been added since the creation of this bookdown). But don't fret, you probably know about these folders already and the most important folders for the dashboard are: **Team_Data**, **Survey_Photos** and **Biodiversity Survey Master**.

8 CHAPTER 2. FILES WITHIN FOLDERS AND FOLDERS WITHIN FILES AND MORE FILES

		Name ▾	Modified ↓ ▾	Modified By ▾	File size ▾	Sharing
		Intern effort	7 September	Will Cresswell	1 item	 Shared
		Biodiversity Survey Master	19 August	Will Cresswell	44 items	 Shared
		Botanic Garden	11 August	Will Cresswell	3 items	 Shared
		Query photos	19 May	Will Cresswell	0 items	 Shared
		Team_Data	16 May	Ane Cloete	10 items	 Shared
		Survey_Photos	16 May	Ane Cloete	470 items	 Shared
		Identification Resources	4 May	Tom Beckett	10 items	 Shared
		Species lists	3 May	Will Cresswell	58 items	 Shared
		Template survey	3 May	Will Cresswell	34 items	 Shared
		How to collect survey data with Qfield file ...	18 May	Leela Stoede	28.2 KB	 Shared
		KPI Biodiversity draft framework.docx	9 May	Will Cresswell	305 KB	 Shared

2.1.1 Biodiversity Survey Master

Here is where all the students upload their biodiversity data. Each student has their own folder labelled with their names and within each folder all the geopackages for each taxa they have collected data on is stored (plus another folder containing backups), e.g.

	Name ▾	Modified ↓ ▾	Modified By ▾	File size ▾	Sharing
	Megan backup	17 May	Megan Gore	39 items	 Shared
	MeganMammals.gpkg	21 September	Ane Cloete	96 KB	 Shared
	MeganButterflies.gpkg	21 September	Ane Cloete	96 KB	 Shared
	MeganBumblebees.gpkg	21 September	Ane Cloete	104 KB	 Shared
	MeganBirds.gpkg	21 September	Ane Cloete	112 KB	 Shared
	MeganMicroMoths.gpkg	21 September	Ane Cloete	96 KB	 Shared
	MeganHoverflies.gpkg	21 September	Ane Cloete	96 KB	 Shared
	MeganGeneric.gpkg	21 September	Ane Cloete	112 KB	 Shared
	MeganMacroMoths.gpkg	21 September	Ane Cloete	100 KB	 Shared
	MeganPlants.gpkg	21 September	Ane Cloete	1.10 MB	 Shared
	MeganTrees.gpkg	21 September	Ane Cloete	180 KB	 Shared

If you don't know what a geopackage is, here's a quick description:

A GeoPackage is like a supercharged file for maps and location data. It's a clever way to bundle up all sorts of info—like where things are on a map, pictures,

and details about those things. Think of it as a digital suitcase for geography. What makes it cool is that it works on different devices and software without any fuss.

2.1.2 Team_Data

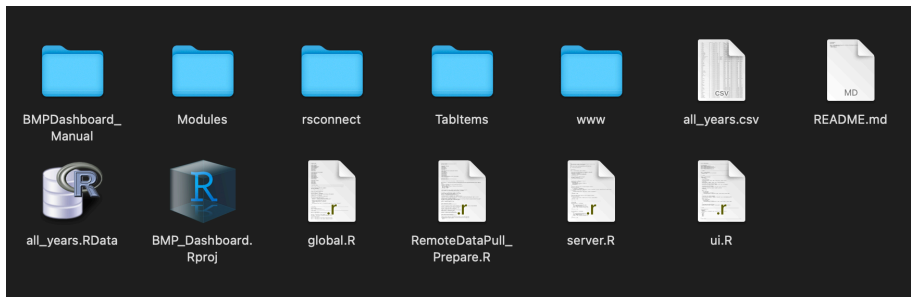
This folder contains photos of each student and a word document containing their “About Me” descriptions used in the dashboard. The nomenclature and file type is important here. The photo must be saved with their name as the file name and they should be jpg’s! Then the word document should be called “student_aboutme” - you can change it if you want, but then you have to change it in the dashboard code as well. Within the document the student descriptions are paragraphs and within the first sentence the student introduces themselves with their names - this is NB. The code will separate the text into paragraphs and then filter for the student by searching for their name. It’s your job to tell everyone to stick to this format and style!

2.1.3 Survey_Photos

All the photos taken while surveying are uploaded here with their unique photo ID as file name. The file name isn’t super important for the dashboard as long as it’s consistent between their records and the uploaded data. Ensure that all the photos are in the same format (i.e. jpg) - if not then there is a way to convert them all in one go which I’ll mention later.

2.2 BMPDashboard

Here is what you should see:



- BMPDashboard_Manual folder
- BMP_Dashboard.Rproj

- RemoteDataPullPrepare.R
- global.R
- server.R
- ui.R
- Modules folder
- rsconnect folder (you'll only see this once you've published the app to shinyapps.io)
- TabItems folder
- www folder
- README.md
- packages.R

If you're familiar with R projects and shiny apps then this should look familiar barring the obvious extras such as the first folder (which contains the code for this bookdown). Let's quickly go over the rest.

2.2.1 RemoteDataPullPrepare.R

This contains the code for pulling the data from OneDrive and preparing it.

2.2.2 global.R, ui.R and server.R

These are the main shiny app files!

2.2.3 TabItems

Contains the ui code for each tab in the dashboard (Tab_About.R, Tab_KPI.R, Tab_Record_Finder.R, Tab_Student_Engagement.R and Tab_Taxa_Explorer.R). I've done this mostly to keep the ui file clean and comprehensible, but it's slightly annoying for work flow as when you make a change here, you have to close the app and then run it again. You could put all the code into the ui script and then if you make changes you only have to reload the app. Up to you!

2.2.4 Modules

Contains the code for a custom valuebox which is used throughout the dashboard which I've turned into a shiny module. A Shiny module is like a neat toolkit for creating specific interactive parts of your app without making a mess of your code. It's like having a mini-app inside your bigger app. So, let's say you want a snazzy chart or a dynamic table—you can build that as a Shiny module. It keeps things organized and clean, making your web development life easier. It's like having building blocks for your website, and each block (or module) does a special job, making your site more interactive and user-friendly.

At the moment I've only converted the valuebox into a module, but there are other things that can be turned into modules too! Like the datatables or the bar graphs. Feel free to play around with this!

2.2.5 **www**

The **www** folder is like the storage room of your shiny app where you keep all the stuff that makes the outside look awesome. In the **www** folder, you put things like images, stylesheets (which control how things look), and JavaScript files (for extra functionality). In our **www** folder you'll find:

- all the survey photos
- a custom css stylesheet (custom.css)
- a copy of the Team_Data folder with additional objects: doc_parts.RData and student_text_sep.RData
- the allyears.RData object (this is the cleaned dataset which the dashboard uses)

2.2.6 README.md

This a plain text file usually written in a simple format called Markdown that contains the description you see on github.

2.2.7 packages.R

This r script contains all code to install and load all the packages required for the dashboard.

Chapter 3

The Data

3.1 Setting up OneDrive

The following instructions are for Mac users. If you are using windows, you should have OneDrive already installed on your computer.

1. Download OneDrive:

- Go to the Mac App Store on your laptop.
- Search for “OneDrive.”
- Click on “Get” or “Install” to download the OneDrive app.

2. Install OneDrive:

- Once downloaded, open your Applications folder and locate the OneDrive app.
- Drag the OneDrive app to your Dock for easier access (optional).

3. Sign In:

- Open the OneDrive app.
- Sign in with your university account.
- Follow the on-screen prompts to set up OneDrive.

4. Choose Folders to Sync:

- Once signed in, you’ll have the option to choose which folders from your OneDrive cloud storage you want to sync with your Mac.
- Select the folders you want or choose to sync everything.

5. Set OneDrive Preferences:

- Click on the OneDrive icon in the Mac menu bar at the top of your screen.

- Click on the three dots (More) and select “Preferences.”
- Here, you can adjust settings like:
 - Starting OneDrive automatically when you sign in to your Mac.
 - Choosing how files are downloaded or uploaded (e.g., over metered networks).
 - Setting up file on-demand (allows you to see all your files without having them downloaded).

6. Accessing Your Files:

- A OneDrive folder will now be present in your Mac’s Finder. This folder will sync with your online OneDrive storage. Any files or folders you add to this folder will be automatically uploaded to the cloud, and any changes you make to files in this folder will be reflected in the cloud version. Read that again! It is important that you know that if you change or delete any files in the OneDrive folder on your laptop it will affect the database online.

7. To Unlink or Sign Out:

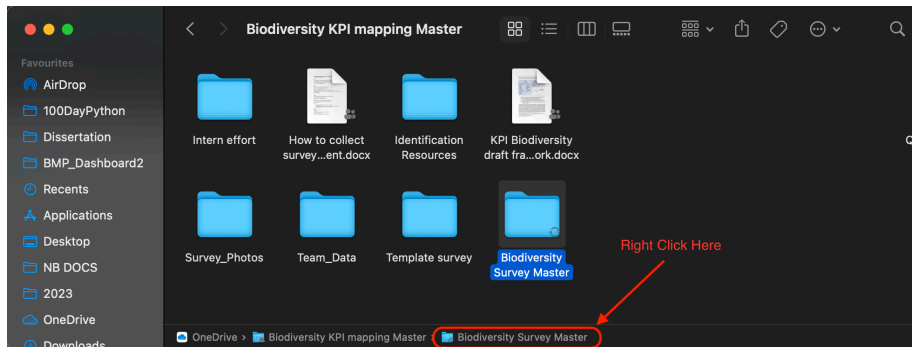
- If you ever wish to unlink your account or sign out, click on the OneDrive icon in the Mac menu bar.
- Click on the three dots (More) and select “Preferences.”
- Go to the “Account” tab and select “Unlink this Mac.”

3.2 Getting the data

First, navigate to the R script called “packages” to check whether you have all the required packages, if you don’t the script will also install and load them for you. Done!

Now, open the script “RemoteDataPull_Prepare”.

The most important first step here is to modify the object “path_to_master” with the filepath to where ever you have the OneDrive folder on your device and specifically to the main survey folder “Biodiversity Survey Master”. If you are using a Mac, navigate to the folder in Finder and then in the bottom panel (see below), right click the folder name and select “Copy”folder name” as Pathname”.



Let's break down the code:

1. Setting the Path to the Master Directory:

```
path_to_master <- "/Users/anecloete/Library/CloudStorage/OneDrive-UniversityofStAndrews/Biodiversity KPI mapping Master/Biodiversity Survey Master"
```

This line sets a variable called `path_to_master` to the path of the main directory where the geopackage files are located. Paste the pathname you copied here.

2. Obtaining All Geopackage Files:

```
all_gpkgs <- list.files(
  path = paste0(path_to_master, "/Biodiversity Survey Master"),
  recursive = TRUE,
  pattern = "\\*.gpkg$",
  full.names = TRUE
)
```

This code lists all files with the “.gpkg” extension in the specified directory and its subdirectories. - The `path` argument specifies where to look. - `recursive` set to `TRUE` means it will look in subdirectories as well. - `pattern` filters for files ending with “.gpkg”. - `full.names` set to `TRUE` ensures the full path of each file is returned, not just its name.

3. Filtering Geopackages with ‘Habitat_Polygon’ in Their Name:

```
survey_polygons <- all_gpkgs[grep("Habitat_Polygon", all_gpkgs)]
```

This filters the `all_gpkgs` vector for filenames that contain the substring “Habitat_Polygon” and assigns the subset to `survey_polygons`.

4. Excluding Geopackages Based on Certain Keywords:

```
pattern <- "Polygon|DESKTOP|PC|LAPTOP|Wills"
all_gpkgs <- all_gpkgs[!grepl(pattern, all_gpkgs)]
```

Here, a pattern is defined to exclude geopackages with certain keywords in their names. The `grepl` function checks for matches, and the `!` operator negates the condition to exclude matches.

5. Exclude a Specific Problematic Geopackage:

```
problematic_gpkg <- paste0(path_to_master, "/Erica/Erica Backup/Ericabutterflies.gpkg")
all_gpkgs <- all_gpkgs[all_gpkgs != problematic_gpkg]
```

This code defines a specific geopackage file path that is “problematic” and then removes this file from the `all_gpkgs` vector.

6. Reading All Geopackages into a List:

```
myfiles <- lapply(all_gpkgs, st_read)
```

The `lapply` function is used here to apply the `st_read` function to each file path in the `all_gpkgs` vector. The result is a list, with each element being the content of a geopackage file stored as an spatial features dataframe. The `st_read` function is part of the `sf` package in R and is used to read spatial data.

In summary, this code: - Sets a path to a master directory. - Lists all geopackage files from this directory and its subdirectories. - Filters and excludes certain geopackages based on keywords or specific filenames. - Reads the content of each remaining geopackage file into a list.

3.3 Data cleaning and preparation

Most of this is self explanatory or is a bit tedious to explain. I would recommended investigating and exploring the data before the cleaning and preparation so that some of the lines make more sense. But here are a few notable things:

- Each student have a backup folder embedded in their folder, these geopackages are removed in line 72
- There is separate geopackage for tree data collected, this is included in the dashboard so is cleaned separately and then joined to the rest.
- The Species column contains the name of the species recorded, whether this is the scientific name, the common name or anything else. So all the naming columns are coalesced into one column called Species.

- Line 135 was necessary because for some reason empty entries in the relevant columns were “ ” and not NA

Some of the code could be a little less sausage making-ish and could be simplified (somethings were added after the fact etc), feel free to make it more efficient!

SUMMARY:

1. Coordinate Transformation:

- Transforms the coordinate reference system of spatial data to standard latitude and longitude (EPSG:4326).

```
# Transform coordinate reference system from WGS 84 / Pseudo-Mercator to standard lat long (EPSG:4326)
taxa_dat <- lapply(myfiles, function(x) st_transform(x, 4326))
```

2. Data Conversions:

- Converts the spatial data frames to regular data frames.
- Names each data frame in the list based on the geopackage filename.
- Removes duplicate data frames based on their names.

```
# Convert spatial data frames to regular data frames
taxa_dat <- lapply(taxa_dat, as.data.frame)

# Name each data frame in the list based on the geopackage filename
names(taxa_dat) <- basename(all_gpkgs)

# Remove any duplicate data frames by name
taxa_dat <- taxa_dat[!duplicated(names(taxa_dat))]
```

3. Tree Entries Package Preparation:

- Modifies columns for a specific geopackage ('Tree Species Entries.gpkg') to ensure compatibility with subsequent operations.
- Cleans and modifies species names.
- Sets certain columns to specific values or NA.
- Filters and selects specific columns.

```
# Modify certain columns for 'Tree Species Entries.gpkg' so that bind_rows works
taxa_dat$`Tree Species Entries.gpkg` <- taxa_dat$`Tree Species Entries.gpkg` %>%
  rename(Species = 1, Date = 2) %>%
  mutate(
    Species = case_when(
      Species %in% c("unknown/other", "Unknown/other") ~ comments.unlisted.species,
```

```

    Species == "Unknown young pine " ~ "Unknown young pine",
    Species == "Willow x10" ~ "Willow",
    TRUE ~ Species
  ),
  taxa = "Vascular Plants",
  Observer = NA,
  photoid = NA,
  Count = NA,
  Other = NA,
  Speciesful = NA
) %>%
filter(Species != "") %>%
select(Species, Date, taxa, Count)

```

4. Date Processing:

- Converts, splits, and extracts components (year, month, day) of the 'Date' column.

```

# Date Processing: Convert, split and extract components of the 'Date' column
taxa_dat <- lapply(taxa_dat, function(df) {

  df$Date <- as.character(df$Date)

  df <- df %>%
    separate(Date, into = c("date", "time"), sep = " (?=[^ ]+)$") %>%
    mutate(
      date = ymd(gsub("/", "-", date)),
      year = year(date),
      month = month(date),
      day = day(date)
    )
  return(df)
})

```

5. Tree Data Extraction:

- Removes the tree data frame from the list and stores it separately.

```

# Extract and remove the tree dataframe from the list
tree_data <- taxa_dat$`Tree Species Entries.gpkg`
taxa_dat$`Tree Species Entries.gpkg` <- NULL

```

6. Combining Data:

- Merges all data frames in the list into a single data frame.

```
# Combine all data frames in the list into a single data frame
taxa_comb <- bind_rows(taxa_dat)
```

7. Data Cleaning and Transformation:

- Performs several operations to clean and transform the combined data, including:
 - Date conversions and modifications.
 - Handling missing values.
 - Excluding certain records.
 - Recoding values in various columns.
 - Selecting and renaming columns.

```
# Clean and transform taxa data for further analysis
taxa_clean <- taxa_comb %>%
  drop_na(taxa) %>%
  as.character(df$Date) %>%
  separate(Date, into = c("date", "time"), sep = " (?=[^ ]+$)") %>%
  mutate(
    date = ymd(gsub("/", "-", date)),
    year = year(date),
    month = month(date),
    day = day(date)
  ) %>%
  filter(!(taxa == "hoverfly" & Observer == "Erica")) %>% # remove Erica hoverfly entries
  unite(collapsed_species, specieslatin:seaweedlatin, sep = ",", na.rm = TRUE) %>%
  mutate_at(vars(Species, collapsed_species), na_if, "") %>%
  mutate(
    Species = coalesce(Species, SpeciesSci, Speciesfull, collapsed_species, species, Other),
    photoid = if_else(is.na(photoid), NA, paste0(photoid, ".jpg")),
    Count = ifelse(is.na(Count), 1, Count),
    taxa = recode(taxa, tree = "Vascular Plants"),
    year = recode(year, `2023` = "2022/2023", `2022` = "2022/2023")
  ) %>%
  filter(!str_detect(Species, "(?i)unknown"), taxa != "bee") %>%
  filter_at(vars(taxa, Observer), all_vars(!is.na(.))) %>%
  mutate(
    taxa = recode(taxa,
      plant = "Vascular Plants",
      bird = "Birds",
      macromoth = "Macromoths",
      micromoth = "Micromoths",
      butterfly = "Butterflies",
```

```

        dragonfly = "Dragonflies",
        hoverfly = "Hoverflies",
        bat = "Bats",
        amphibian = "Amphibians",
        reptileamphibian = "Amphibians",
        bumblebee = "Bumblebee",
        mammal = "Mammals",
        ladybird = "Ladybirds",
        tree = "Vascular Plants"),
  year = recode(year,
    "2023" = "2022/2023",
    "2022" = "2022/2023"),
  Observer = recode(Observer,
    Other1 = "Cori")) %>%
select(Species, SpeciesSci, Count, date, Observer, taxa, photoid, geometry, year, day)
rename(
  Date = date,
  Taxa = taxa,
  PhotoID = photoid
)

```

8. Merging Tree Data with Cleaned Data:

- Prepares the tree data to be merged with the cleaned data.
- Merges both datasets.

```

# tree data prep for bind
tree_data <- tree_data %>%
  mutate(
    year = ifelse(year == "2023" | year == "2022", "2022/2023", year),
    Count = 1 # add column count
  ) %>%
  rename(
    Date = date,
    Taxa = taxa
  )

# Merge tree data with the cleaned data
all_years <- bind_rows(taxa_clean, tree_data)

```

9. Geometry Processing:

- Splits the geometry column into separate latitude and longitude columns.

```
# split into lat and long for mapping
all_years <- all_years %>% mutate(long = unlist(map(geometry,1)),
                                   lat = unlist(map(geometry,2)))
```

10. Saving the Resulting Data:

- Saves the final data frame both as an RData object and as a CSV file.

```
# Save the resulting dataframe
save(all_years, file = "all_years.RData")
write.csv(all_years, file = "all_years.csv")
```

3.4 Pulling photos and documents from OneDrive

3.4.1 Photos

1. Reading Team Photos from a Directory:

```
# read in all photo files within Survey_Photos in OneDrive Folder
all_Tphotos <- list.files(
  path = paste0(path_to_master, "/Team_data"),
  recursive = TRUE,
  pattern = "\\\\.jpg$",
  full.names = FALSE
)
```

This is the same as the first few lines of code, except that the path is constructed by appending “/Team_data” to the `path_to_master`.

The result is stored in `all_Tphotos`, which will be a vector of filenames (with “.jpg” extension) from the specified directory and its subdirectories.

2. Constructing a Destination Path for Team Photos:

```
www_Tfolder_path <- file.path(getwd(), "www/Team_Data")
```

This line constructs a path by combining the current working directory (obtained using `getwd()`) with “www/Team_Data”.

3. Copying the Team Photos:

```
photo_move <- file.copy(  
  from = file.path(paste0(path_to_master, "/Team_data"), all_Tphotos),  
  to = file.path(paste(www_Tfolder_path), all_Tphotos)  
)
```

Here, `file.copy` is used to copy files. The `from` argument constructs the full paths of the source files, and the `to` argument constructs the full paths of the destination. The photos are copied from the source directory to the destination.

The same process as above is then done for the Survey Photos.

In summary, the code is designed to:

- Read all “.jpg” files from “Team_data” and “Survey_Photos” directories (including subdirectories).
- Copy those files to two new destinations under the “www” folder in the current working directory.

3.4.2 Documents

The code for downloading and saving the students’ about me descriptions is very similar to the process above. The only addition is that the text of the “student_aboutme” word document is then extracted and saved into an object called “student_text”. Then the text is split into paragraphs.

3.5 Things to think about

How will the next years data be included? In the same folder? New QGIS folder?

Chapter 4

Footnotes and citations

4.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

4.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2023] (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [Xie, 2015] (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 5

Blocks

5.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (5.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (5.1).

5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 5.1.

Theorem 5.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

5.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Chapter 6

Next Steps

Here is a list of things that need to be done:

6.1 Different years different data:

Currently, the dashboard works a dataset containing data for multiple years - I created a fake dataset for the next academic year as a subset of the 2023 data. The data is combined into the `all_years.RData`. It might be better to have separate datasets and store them in a list, its up to you. The next point will influence this too.

6.2 Data download and clean automation:

The pulling and preparing of the data should be made automatic, i.e. everyday during the survey season the data should be pulled from OneDrive, cleaned and uploaded to the dashboard. However, as more data is collected it will probably be better to upload the data to OneDrive (i.e. not host it on your laptop or shiny server) and have the dashboard pull the data from OneDrive. This will also affect how the dashboard accesses separate years data.

6.3 Survey Photos:

The same also applies to the photos. It might be better to pull the photos directly from OneDrive in the app. I'm not sure how to do this and maybe there's a different way to store and retrieve the photos.

Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL <http://yihui.org/knitr/>. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2023. URL <https://CRAN.R-project.org/package=bookdown>. R package version 0.34.