

Coding style guide

R

DIRECTORIES

- If data and output folders are located within the project folder, set their directories at the top of the code and do this dynamically using `getwd() + file.path()`
- If these folders are located outside the project folder (e.g., they are too big to live on GitHub), then just be sure to provide the file path at the top

DOCUMENTATION

- Top of code should describe what the code does
- Use a space after “#” (Mac: Command + Shift + C; Windows: Ctrl + Shift + C)
- For “internal” or “personal” functions (e.g., not part of a package that will be released by the lab), include name and email at the top – just good practice and assigns “authorship” to you (when you share code with people they know who to contact with questions)
- While code is in development, use “# FIX IT ” to flag areas of the code that need more attention
- It’s OK to break up long sections of piped (`%>%`) code with documentation (e.g., creation of `output_df` in `commod_commod_matching.R`)

CODE

- Assignment arrows should point left, not right
- Pay attention to indentation (convention is two spaces), use package “styler” (can be accessed as an Add-In in RStudio) and use default “tidyverse style” to clean indentation
- Use spaces around `=`, `<-`, and `%>%`

Example

Avoid:

```
iris%>% dplyr::summarize(max_petal=max(Petal.Width))->results
```

Prefer:

```
Results <- iris %>% dplyr::summarize(max_petal = max(Petal.Width))
```

NAMING OBJECTS

- All lower case
- Prefer underscores (`seafood_species`) over camel case (`seafoodSpecies`) for all object names
- For data frame columns, can use “.” or “_” as a sep in column names, just be consistent
- Function names should begin with a verb (e.g., `clean_cols`, `tidy_data`, `calc_distance`), be concise, and use underscores to separate words

- Object/Variable names should be nouns and descriptive (e.g., `df_clean` or `clean_df`)

Example:

Avoid: `buildFishMatrix`

Prefer: `build_fish_matrix`

Avoid: naming objects with different capitalization (e.g., `HS_codes`, `hs_Codes`, `hsCodes`)

Prefer: `hs_codes_raw`, `hs_codes_filtered`, `hs_codes_cleaned`, etc.

NAMING OUTPUT FILES

- If outputs are going to be different with each run (e.g., running different analyses), start filename of outputs with a date
- Follow international date convention (YYYY-MM-DD) (Can use `Sys.Date()` for this)
- Use delimiters with intention (e.g., “`_`” vs “`-`”) so they are machine readable
- Use “`-`” for spaces and “`_`” for separating metadata

Example:

Prefer: `YYYY-MM-DD_name-of-file_i_alpha-4-cutoff.csv`

Avoid: `NameOfFileResults-DD-MM-YYYY-i_alpha_4_cutoff.csv`

PACKAGES

Each .R file in the R folder should be a separate function