Module 4 – lesson 03: Create an R Markdown Template and Share it

Script

For this lesson, let’s create a new Github repository and RStudio project. Create a Github repository called “WCtemplateSimple” with a readme file. Type in a description such as “Weather Check - Simple R Markdown Template”. Then in RStudio create a new project with Version Control via Git for this repository. We’re going to use this project and repository to create a new R markdown template for creating a report using the Weather Check dataset from the fivethirtyeight R package.

To build our template, let’s reuse some of the files we’ve already created. Get a copy of the “steakArticleParamsList.Rmd” R markdown file with the parameters list for the various United States regions we created in Module 3 for the steak\_survey dataset from the fivethirtyeight package. Put a copy of this R markdown file “steakArticleParamsList.Rmd” in the directory for this “WCtemplateSimple” project.

In RStudio, let’s create an R script to briefly explore the weather\_check dataset in the fivethiryeight R package. Go to File/New File/R Script. To explore the dataset, load the fivethirtyeight R package and load the weather\_check dataset with the following code

library(fivethirtyeight)

data("weather\_check", package="fivethirtyeight")

We’ll next use the names() function to explore the variables in this dataset.

names(weather\_check)

We can also get additional information through the R help window – click on the Packages TAB and click on the link for the fivethirtyeight package and then scroll to the bottom to see the link for the weather\_check dataset. You’ll notice that many of the variables listed for this dataset are the same as those we explored earlier in the steak\_survey dataset – specifically, the demographics for age, female, hhold\_income and region. This means we can reuse some of the same R code we had before with a few minor changes.

Given that we used the United States regions as the variable for the document parameter in the steak\_survey it would be nice if we could do that again here for the weather\_check dataset. Let’s verify that the regions in both datasets have the same choices. To do that let’s use the table() function to get a quick summary of the region choices in both datasets.

table(weather\_check$region)

table(steak\_survey$region)

This is great – both datasets have the same choices spelled exactly the same in both datasets. This means we can reuse the YAML header and parameter options from the steak\_survey template for region in a new R markdown template for creating reports on the weather\_check dataset.

In RStudio, open the “steakArticleParamsList.Rmd” R markdown file. Then save the file as “wcArticle.Rmd”. Let’s make some changes. Refer to the read ahead supplementary materials to assist you with these changes.

In the YAML header change the title to

“Weather Information Preferences in the United States”

The rest of the YAML header can stay the same. If you want, feel free to tweak the html\_document theme if you like.

We also need to update the first R code chunk to switch out the dataset from steak\_survey to weather\_check. And the na.omit() was removed for this dataset.

[ON COMPUTER – type in this code]

# load steak\_survey dataset

data("weather\_check", package="fivethirtyeight")

# create subset, no missing NAs

# pick a specific region

sdat <- weather\_check %>%

filter(region==params$region)

It is worth noting that we could make other changes, but let’s try to keep as much of the code the same as possible so we can be efficient about reusing what we did previously.

We’ll keep the sections headers mostly the same. In the background section, let’s update the text to include details on the original 538.com article on “Where People Go To Check The Weather” Update the Background section using the text provided in the read ahead supplementary materials.

[COPY TO RStudio for COMPUTER DEMO]

In April 2015, Walt Hickey at [538.com](http://fivethirtyeight.com/) published an article entitled ["Where People Go To Check The Weather"](<https://fivethirtyeight.com/features/weather-forecast-news-app-habits/>). This article utilized survey data collected from 928 people which asked questions on if they typically check the weather on a daily basis and if so, what source they used to check the weather. Additional follow-up questions were asked about which specific websites or apps they used and if they had a smartwatch if they would check the weather on that smartwatch.

For the Purpose section put in the updated section provided in the read ahead supplementary materials. Again notice that only minor changes were made – mostly swapping out references to the steak survey dataset for the weather check dataset.

[COPY over for COMPUTER DEMO]

The "weather check" dataset that was used for the ["Where People Go To Check The Weather"](https://fivethirtyeight.com/features/weather-forecast-news-app-habits/) article, is available through the R package [`fivethirtyeight`](https://cran.r-project.org/web/packages/fivethirtyeight/index.html). Using this dataset, this article summarizes the weather source preferences for the `r params$region` Region of the United States by gender.

For our summary presented here, a subset of the original weather check dataset was extracted to only include responses with no missing data and only included respondents from the `r params$region` region of the US. The data subset used for our summary had `r nrow(sdat)` respondents.

Let’s remove the next section which summarized the demographics in separate tables and instead just keep the section where the summary tables were merged. Instead, let’s update the previously optional section as follows – refer to the content provided in the read ahead supplementary materials. You’ll notice that the changes were minor. It is worth noting that the weather check dataset did not have education data on the respondents. Instead we’ll use the female variable to do a summary of gender in the demographics table.

[COPY over for COMPUTER DEMO]

## Demographics of Survey Respondents in a Merged Table

This section provides demographic summary tables for the `r nrow(sdat)` survey respondents.

```{r}

# create table summary for ages

tb1 <- sdat %>%

select(age) %>%

table() %>%

prop.table()\*100

# create table summary for gender

tb2 <- sdat %>%

select(female) %>%

table() %>%

prop.table()\*100

# create table summary for income

tb3 <- sdat %>%

select(hhold\_income) %>%

table() %>%

prop.table()\*100

# convert all tables to data frames

tb1df <- as.data.frame(tb1)

tb2df <- as.data.frame(tb2)

tb3df <- as.data.frame(tb3)

# merge 1st 2 data frames together

mtb <- merge(data.frame(tb1df, row.names = NULL),

data.frame(tb2df, row.names = NULL),

by=0, all=TRUE)[-1]

# merge result with 3rd data frame

mtb2 <- merge(data.frame(mtb, row.names = NULL),

data.frame(tb3df, row.names = NULL),

by=0, all=TRUE)[-1]

# use the final data frame

# make into a table with kable

# add styling with kableExtra

# add header with labels spanning 2 columns each

mtb2 %>%

knitr::kable(format="html",

col.names=c("Category","%",

"Category","%",

"Category","%"),

digits=2,

caption="Demographics of Survey Respondents") %>%

kableExtra::kable\_styling(c("striped","bordered"),

full\_width=FALSE) %>%

add\_header\_above(c("Ages"=2,"Gender"=2,"Income"=2))

```

For the final section, make minor changes to the section header and text to change the steak preferences to weather source. Copy over the updated text and R code from the read ahead supplementary materials. In the ggplot code a line was added for coord\_flip() which makes a horizontal bar chart instead of the vertical bar chart we had previously. This change was made to allow for the longer response choices for weather source. The stringr library was also loaded from the tidyverse package so the str\_wrap function could be used to wrap the long response choices to fit better within the clustered bar chart.

[COPY over for COMPUTER DEMO]

## Weather Source Preference by Gender

Finally, here is a breakdown of the `r nrow(sdat)` survey respondents for the `r params$region` region of the US on how they check their daily weather by gender.

```{r}

library(stringr)

ggplot(sdat,

aes(x = weather\_source, fill = female)) +

geom\_bar(position="dodge",colour="black") +

scale\_x\_discrete(labels = function(x) str\_wrap(x, width = 20)) +

ggtitle(paste0("Weather Source by Gender: ",

params$region," Region")) +

xlab("Weather Source Preference") +

ylab("Number of Respondents") +

scale\_fill\_manual(values=c("skyblue","palevioletred"),

name="Gender",

breaks=c(FALSE,TRUE),

labels=c("Male", "Female")) +

coord\_flip() +

theme\_fivethirtyeight()

```

Click Save and KNIT with Parameters – choose a region and Click KNIT to see the result. Test the template with different regions and check to see how everything looks. Admittedly there is still a little more tweaking that could be done to fine tune the table and figure, but we can work well with what we have so far.

Now let’s go ahead and back everything up to your Github account.

Open Git Bash and make sure you are in the correct directory:

C:\RepTemplates\WCtemplateSimple

Once in that directory, type in the following 4 Git commands to check the status of your local files compared to your Github cloud repository; add or stage the modified files; commit your changes; and then push the changes to your Github cloud repository.

git status

git add .

git commit –m “add template example documents”

git push

Now go to your Github repository, refresh if necessary and see your newly committed files.

From this point, if you would like to share your template with others, you could simply send them a link to your Github repository and they could download your whole Github repository as a ZIP file or simply download the “wcArticle.Rmd” file. You could also send the “wcArticle.Rmd” file to someone as an email attachment or via another document sharing resource. It is worth noting that if someone else also has a Github account they could simply “fork” a copy of your repository to their own account.

As a quick example, suppose you wanted to make a copy of the tufte package for your own uses. Log into your Github account. Then go to the Github repository for the tufte package at <https://github.com/rstudio/tufte>. A the top right there is an icon that looks like a Y next to the word Fork. Click on this icon and Github will make a copy of the tufte package repository into your own Github account. This process will take a minute to complete. Once this process completes, you can now work with your copy of the tufte package repository like you would work with any other of your Github repositories. If you wanted to customize the tufte template included with this package, you could make changes and use the updated package for your own purposes. It is worth noting that this will NOT affect the original tufte package. There are ways to collaborate by forking someone else repository, making changes and then proposing to the original developer that they merge your changes back into their original repository. This workflow is beyond the scope of this course, but you can learn a lot more on this topic in the Github help pages for “fork a repo” at <https://help.github.com/articles/fork-a-repo/>

In the next lesson, you’re going to take the weather check template we just created and develop your own R package, install it and recall your template using the RStudio interface for creating a new R markdown template file.