

homework_01

David Coomes

1/12/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```

# make a stupid directory
stupid_dir <- file.path(Sys.getenv("TEMP"), "delete_me_later")

if(!dir.exists(stupid_dir)){
  dir.create(path = stupid_dir)
}

# move into the directory
setwd(stupid_dir)

# a data frame to write to a file
x <- data.frame(foo = "bar")

# make some stupidly named files, Dec 16, 2021
write.csv(x = x, file = file.path(stupid_dir, "file 121621.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file 12_16_21.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file 12-16-21.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file 16-12-21.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file_121621.csv"), quote = FALSE, row.names = FALSE)

# make some stupidly named files, Dec 17, 2021
write.csv(x = x, file = file.path(stupid_dir, "file 121721.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file 12_17_21.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file 12-17-21.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file 17-12-21.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file_121721.csv"), quote = FALSE, row.names = FALSE)

# write two logically named files
write.csv(x = x, file = file.path(stupid_dir, "file_20211216.csv"), quote = FALSE, row.names = FALSE)
write.csv(x = x, file = file.path(stupid_dir, "file_20211217.csv"), quote = FALSE, row.names = FALSE)

message("which file is newer?")
write.csv(list.files(stupid_dir), quote = FALSE, row.names = FALSE)
message("which file is older?")
write.csv(sort(list.files(stupid_dir), decreasing = TRUE), quote = FALSE, row.names = FALSE)

message(paste("You should be convinced that yyyyymmdd is a better way to name files by date.\nYour reward"))

file_name <- c("Instrument_v.20 - bilingual.docx"
, "Master_Durham.csv"
, "Master_Durham.xlsx"
, "Durham emails_Nov 30.xlsx"
, "NJ emails_Nov 30.xlsx"
, "Durham tasks_Nov 30.xlsx"
, "Master_Durham_v2.csv"
, "Master_Durham_v2.xlsx"
, "NJ emails_December 7.xlsx"
, "NJ phone numbers_Nov 30.xlsx"
, "Pitt phone numbers_December 7.xlsx"
, "Pitt phone numbers_Nov 30.xlsx"
, "Fraud and quota report_v1.xlsx"
, "Fraud and quota report_v2.xlsx"
, "Rutgers_people_to_pay_12_11_2020.xlsx"
, "Durham emails_Dec 14.xlsx"

```

```

,"Durham phone numbers_Dec 14.xlsx"
,"Export_1215.7z"
,"NJ emails_December 14.xlsx"
,"NJ phone numbers_December 14.xlsx"
,"VR_Dec1420_Durham.xlsx"
,"VR_NJ_Dec1420.xlsx"
,"transfer_0105.zip"
,"Fraud and quota report_v3.xlsx"
,"Fraud and quota report_v4.xlsx"
,"test1.dta")

date <- c("2020-10-13",
          "2020-11-30",
          "2020-11-30",
          "2020-12-04",
          "2020-12-04", "2020-12-07", "2020-12-07",
          "2020-12-07",
          "2020-12-07",
          "2020-12-07",
          "2020-12-07",
          "2020-12-11",
          "2020-12-11",
          "2020-12-11",
          "2020-12-15",
          "2020-12-15",
          "2020-12-15",
          "2020-12-15",
          "2020-12-15",
          "2020-12-15",
          "2020-12-15",
          "2021-01-05",
          "2021-01-07",
          "2021-01-07",
          "2021-01-08")

file_name <- c("Instrument_v.20 - bilingual.docx"
              ,"Master_Durham.csv"
              ,"Master_Durham.xlsx"
              ,"Durham emails_Nov 30.xlsx"
              ,"NJ emails_Nov 30.xlsx"
              ,"Durham tasks_Nov 30.xlsx"
              ,"Master_Durham_v2.csv"
              ,"Master_Durham_v2.xlsx"
              ,"NJ emails_December 7.xlsx"
              ,"NJ phone numbers_Nov 30.xlsx"
              ,"Pitt phone numbers_December 7.xlsx"

```

```

, "Pitt phone numbers_Nov 30.xlsx"
, "Fraud and quota report_v1.xlsx"
, "Fraud and quota report_v2.xlsx"
, "Rutgers_people_to_pay_12_11_2020.xlsx"
, "Durham emails_Dec 14.xlsx"
, "Durham phone numbers_Dec 14.xlsx"
, "Export_1215.7z"
, "NJ emails_December 14.xlsx"
, "NJ phone numbers_December 14.xlsx"
, "VR_Dec1420_Durham.xlsx"
, "VR_NJ_Dec1420.xlsx"
, "transfer_0105.zip"
, "Fraud and quota report_v3.xlsx"
, "Fraud and quota report_v4.xlsx"
, "test1.dta")

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.2.1      v purrr 0.3.3
## v tibble 2.1.3       v dplyr 1.0.2
## v tidyr 1.0.0        v stringr 1.4.0
## v readr 1.3.1        v forcats 0.4.0

## Warning: package 'dplyr' was built under R version 3.6.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

table_1 <- tibble(date, file_name)

new_file_name <- c("instrument_bilingual_20201013.docx", "master_durham_20201130.csv", "master_durham_20201130.xlsx", "master_durham_20201207.csv", "master_durham_20201207.xlsx", "pitt_phone_numbers_20201207.xlsx", "pitt_phone_numbers_20201130.xlsx", "fraud_quota_report_20201211.xlsx", "fraud_quota_report_v2_20201211.xlsx", "rutgers_people_to_pay_12_11_2020.xlsx", "durham_phone_numbers_20201214.xlsx", "export_20201215.7z", "nj_emails_20201214.xlsx", "vr_nj_20201214.xlsx", "transfer_20210105.zip", "fraud_quota_report_20210107.xlsx", "test1.dta")

table_2 <- bind_cols(table_1, new_file_name) %>%
  rename(new_file_name = "...3")

## New names:
## * NA -> ...3

library(kableExtra)

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
## group_rows

kable(table_2) %>% kable_styling(bootstrap_options = c("striped", "condensed"), full_width=TRUE)

```

date	file_name	new_file_name
2020-10-13	Instrument_v.20 - bilingual.docx	instrument_bilingual_20201013.docx
2020-11-30	Master_Durham.csv	master_durham_20201130.csv
2020-11-30	Master_Durham.xlsx	master_durham_20201130.xlsx
2020-12-04	Durham emails_Nov 30.xlsx	durham_emails_20201130.xlsx
2020-12-04	NJ emails_Nov 30.xlsx	nj_emails_20201130.xlsx
2020-12-07	Durham tasks_Nov 30.xlsx	durham_tasks_20201130.xlsx
2020-12-07	Master_Durham_v2.csv	master_durham_20201207.csv
2020-12-07	Master_Durham_v2.xlsx	master_durham_20201207.xlsx
2020-12-07	NJ emails_December 7.xlsx	nj_emails_20201207.xlsx
2020-12-07	NJ phone numbers_Nov 30.xlsx	nj_phone_numbers_20201130.xlsx
2020-12-07	Pitt phone numbers_December 7.xlsx	pitt_phone_numbers_20201207.xlsx
2020-12-07	Pitt phone numbers_Nov 30.xlsx	pitt_phone_numbers_20201130.xlsx
2020-12-11	Fraud and quota report_v1.xlsx	fraud_quota_report_20201211.xlsx
2020-12-11	Fraud and quota report_v2.xlsx	fraud_quota_report_v2_20201211.xlsx
2020-12-11	Rutgers people to pay 12_11_2020.xlsx	rutgers_people_topay_20201211.xlsx
2020-12-15	Durham emails_Dec 14.xlsx	durham_emails_20201214.xlsx
2020-12-15	Durham phone numbers_Dec 14.xlsx	durham_phone_numbers_20201214.xlsx
2020-12-15	Export_1215.7z	export_20201215.7z
2020-12-15	NJ emails_December 14.xlsx	nj_emails_20201214.xlsx
2020-12-15	NJ phone numbers_December 14.xlsx	nj_phone_numbers_20201214.xlsx
2020-12-15	VR_Dec1420_Durham.xlsx	vr_durham_20201214.xlsx
2020-12-15	VR_NJ_Dec1420.xlsx	vr_nj_20201214.xlsx
2021-01-05	transfer_0105.zip	transfer_20210105.zip
2021-01-07	Fraud and quota report_v3.xlsx	fraud_quota_report_20210107.xlsx
2021-01-07	Fraud and quota report_v4.xlsx	fraud_quota_report_v2_20210107.xlsx
2021-01-08	test1.dta	test_20210108.dta