

Reproducible Reporting with RMarkdown (R3)

Mar. 2022



01

...

Preface

02

...

What's RMarkdown?

Agenda

03

...

How to make reports with RMarkdown?

04

...

Report Exapmles

01

...

Preface



Motivation for Reporting Task

Robust Research:

- “Robust research is about doing small things that stack the deck in your favor to prevent mistakes.”
—Vince Buffalo, author of Bioinformatics Data Skills (2015)

Reproducible Research:

- Reproducible research can be repeated by other researchers with the same results

Interactive Reporting:

- Apart from paper or PDF reports, allowing users to interact with the report allows them to ask questions about the data itself



Requirements for Reporting

5

Distributable:

- Self-contained report that makes it easier to get feedbacks

Viewable to Anyone:

- Viewable to non-engineers in order to eliminate information asymmetry
e.g.) In most cases, JupyterLab can only be accessed by engineers

Automate Reporting:

- Reporting task should be done at the same time as analysis task
- Preliminary analytical design becomes more important



02

...

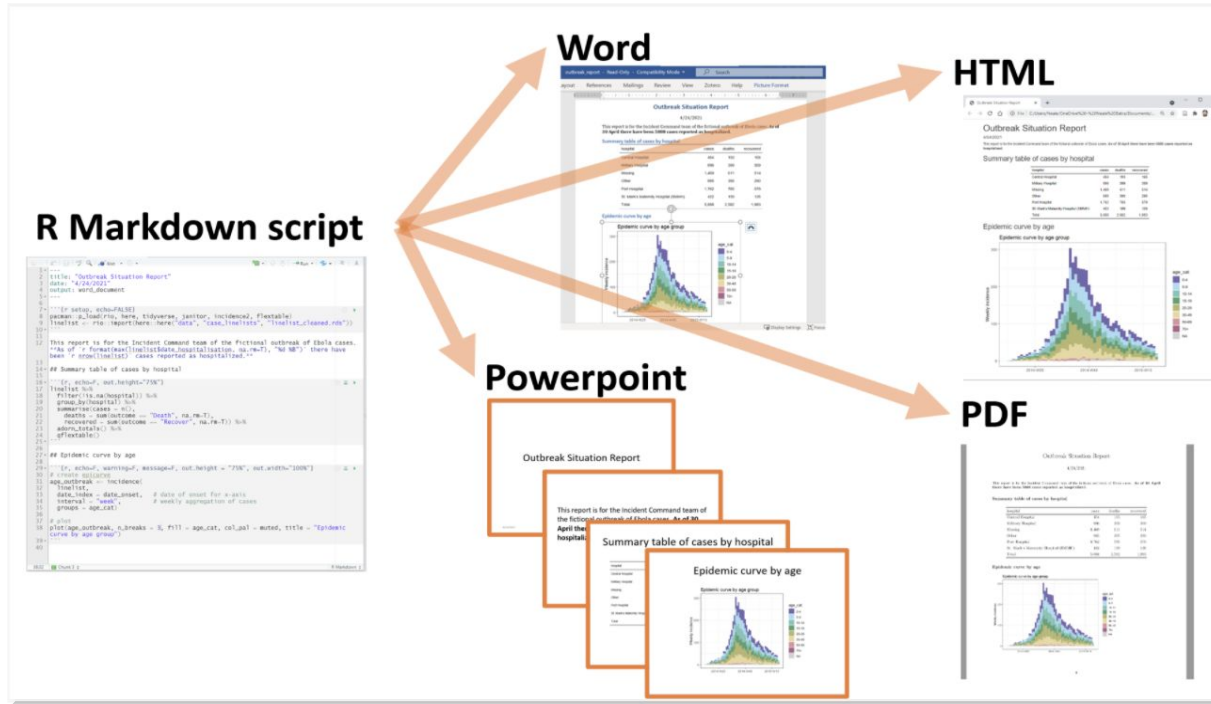
What's RMarkdown?





What's RMarkdown?

RMarkdown is a widely-used tool for creating automated, reproducible, and share-worthy outputs, such as reports (html, pdf, docx, ...)



[Source]



What's R Markdown?

R Markdown is a widely-used tool for creating automated, reproducible, and share-worthy outputs, such as reports (html, pdf, docx, ...)

R Markdown script

```
1 title: "Outbreak Situation Report"
2 date: "4/24/2021"
3 output: word_document
4
5
6
7 ---[r setup, echo=FALSE]
8 pacman::p_load(ro, here, tidyverse, janitor, incidence2, flextable)
9 line1st <- rio::import(here("data"), "case_line1sts", "line1st_cleaned.rds")
10
11
12 This report is for the Incident Command team of the fictional outbreak of Ebola cases.
13 **As of `r format(max(line1st$date_hospitalisation, na.rm=T), "%d %B")` there have
14 been `r nrow(line1st)` cases reported as hospitalized.**
15
16 ## Summary table of cases by hospital
17
18 ---[r, echo=F, out.height="75%"]
19 line1st %>%
20 filter(!is.na(hospital)) %>%
21 group_by(hospital) %>%
22 summarise(cases = n(),
23           deaths = sum(outcome == "Death", na.rm=T),
24           recovered = sum(outcome == "Recover", na.rm=T)) %>%
25 arrange(desc(cases))
26
27 ## Epidemic curve by age
28
29 ---[r, echo=F, warning=F, message=F, out.height = "75%", out.width="100%"]
30 # create epidemic curve
31 age_outbreak <- incidence(
32   line1st,
33   date_index = date_onset, # date of onset for x-axis
34   interval = "week", # weekly aggregation of cases
35   groups = age_cat)
36
37 # plot
38 plot(age_outbreak, n_breaks = 3, fill = age_cat, col_pal = muted, title = "Epidemic
39 curve by age group")
40
```

Annotations for R Markdown script:

- YAML sets title, date, and output type
- Code chunk loads packages and data
- Text and in-line code
- Code chunk makes table
- ## Headings
- Code chunk makes plot

Output (e.g. Word document)

Annotations for Output:

- Text and in-line code
- Code chunk makes table
- ## Headings
- Code chunk makes plot

hospital	cases	deaths	recovered
Central Hospital	454	193	165
Military Hospital	896	399	309
Missing	1,409	611	514
Other	885	395	290
Port Hospital	1,762	765	579
St Mark's Maternity Hospital (SMBH)	422	199	126
Total	5,888	2,562	1,963

[Source]



Benefits of RMarkdown

Easy to Deliver:

- We can distribute html reports. Other formats (pdf, docx) are available
- Can be attached to ESA (up to 10MB by default)

Easy to Reproduce:

- Easy to recreate the report if the report data is updated
- We can easily create reports if know Markdown

Rich Expression:

- Interactive report (only html format)
- Ingenuities that can aggregate information, such as Tab / Table display

03

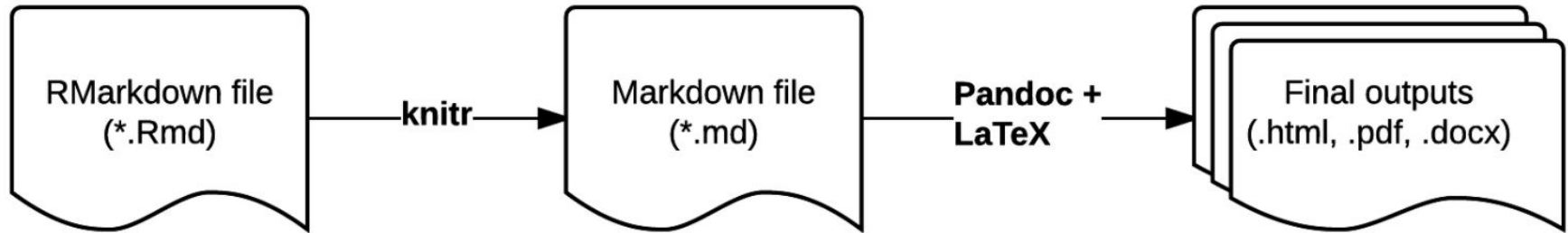


How to make reports with RMarkdown



RMarkdown Workflow: [\[Link\]](#)

- RMarkdown is an enhanced version of Markdown that lets you embed R code into the docs
- `.Rmd` → `.md` → Document(`.html/.pdf/.docx`)



Creation example by R

- Prepare 2 type of files for report contents(.Rmd) and rendering(.R)

Report contents(.Rmd)

```
# Let's embed some R code
```${r}
library(dplyr)
library(readr)
gm <- read_csv('data/gapminder.csv')
````
```

The mean life expectancy is
`\${r} mean(gm\$lifeExp)` years.

The years surveyed in this data include:
`\${r} unique(gm\$year)`.

Rendering(.R)

```
library(rmarkdown)

rmarkdown::render(
  input=input_file_Rmd,
  output_format="html_document",
  output_file=output_file,
  encoding='utf-8',
  quiet=F)
```

Call R from Python

- We can easily call R from Python

```
import os
os.system("Rscript ./render.R --arg1=xxx --arg2=yyy, ...")
```

- We can also use R with Google Colaboratory, just adding below

```
[ ] # activate r magic
    %load_ext rpy2.ipynon
```

```
[ ] %%R
    Sys.setenv(TZ="Asia/Tokyo")
    install.packages('argparse')
    install.packages('rmdformats')
    install.packages('formattable')
    install.packages('DT')
```

04

...

Report Examples

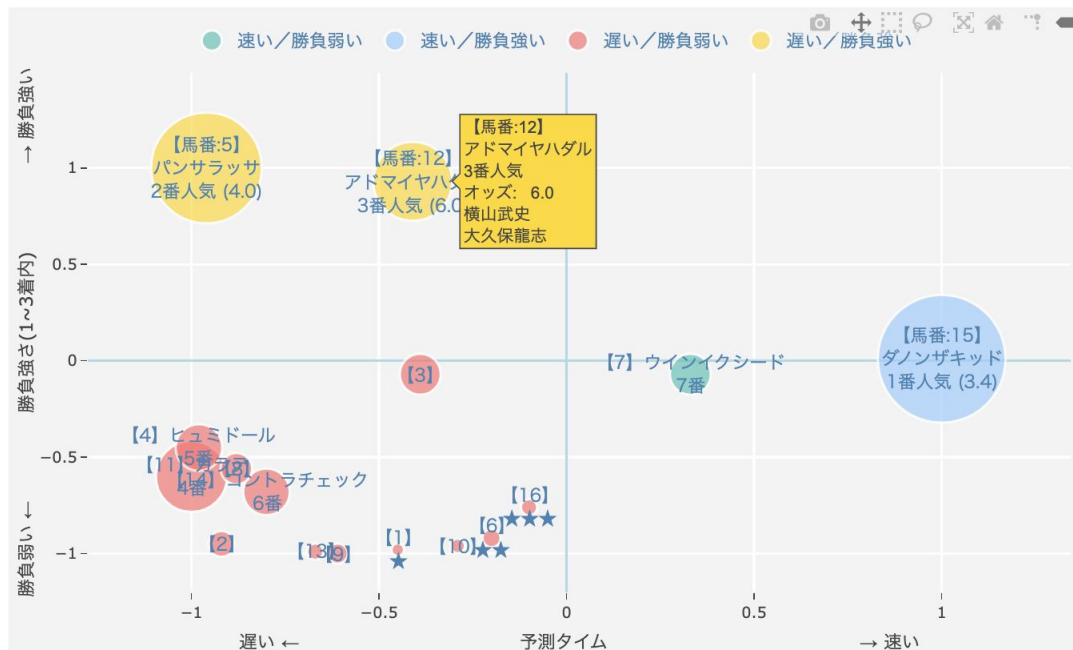


Ex.1) Horse Racing Prediction(1/4): [\[Link\]](#)

15

Visualize the prediction scores of speed (x-axis) and clutch (y-axis)

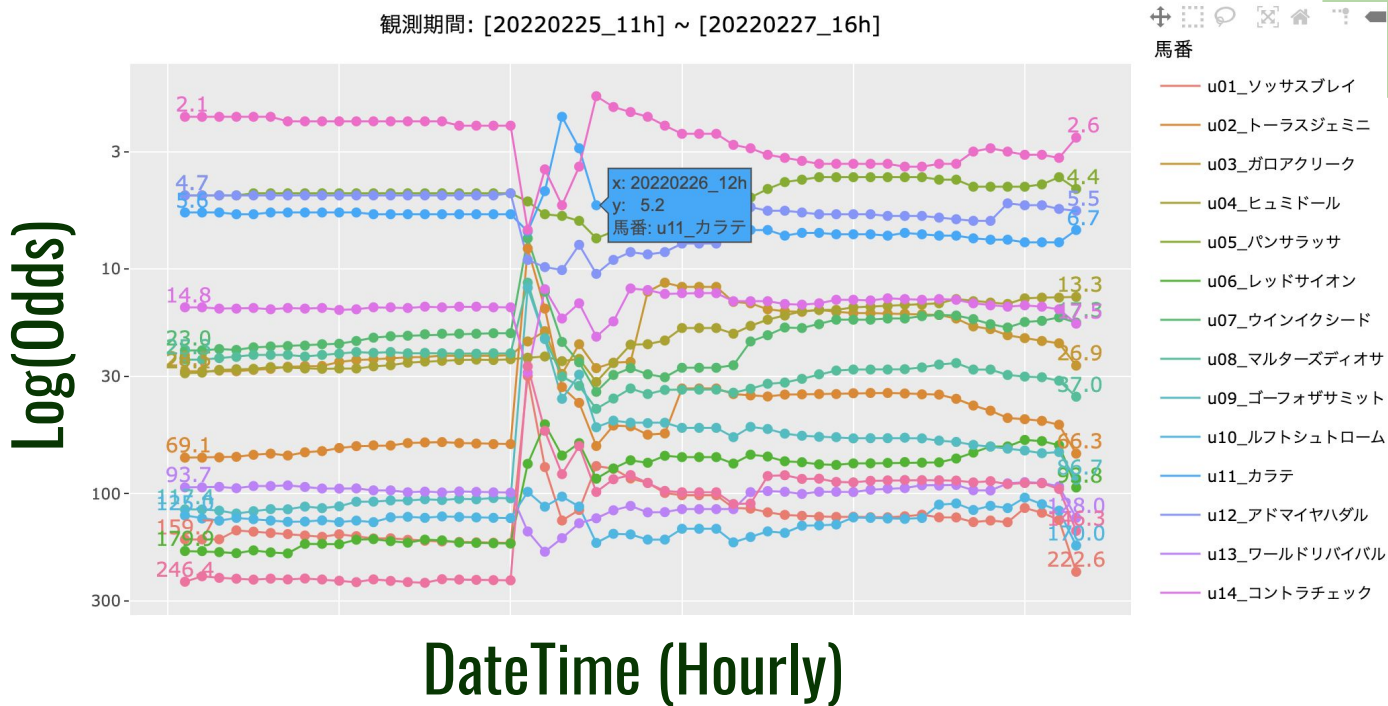
Prediction Chart (Speed/Clutch)



Ex.1) Horse Racing Prediction(2/4): [\[Link\]](#)

Odds information will be updated hourly until just before the race

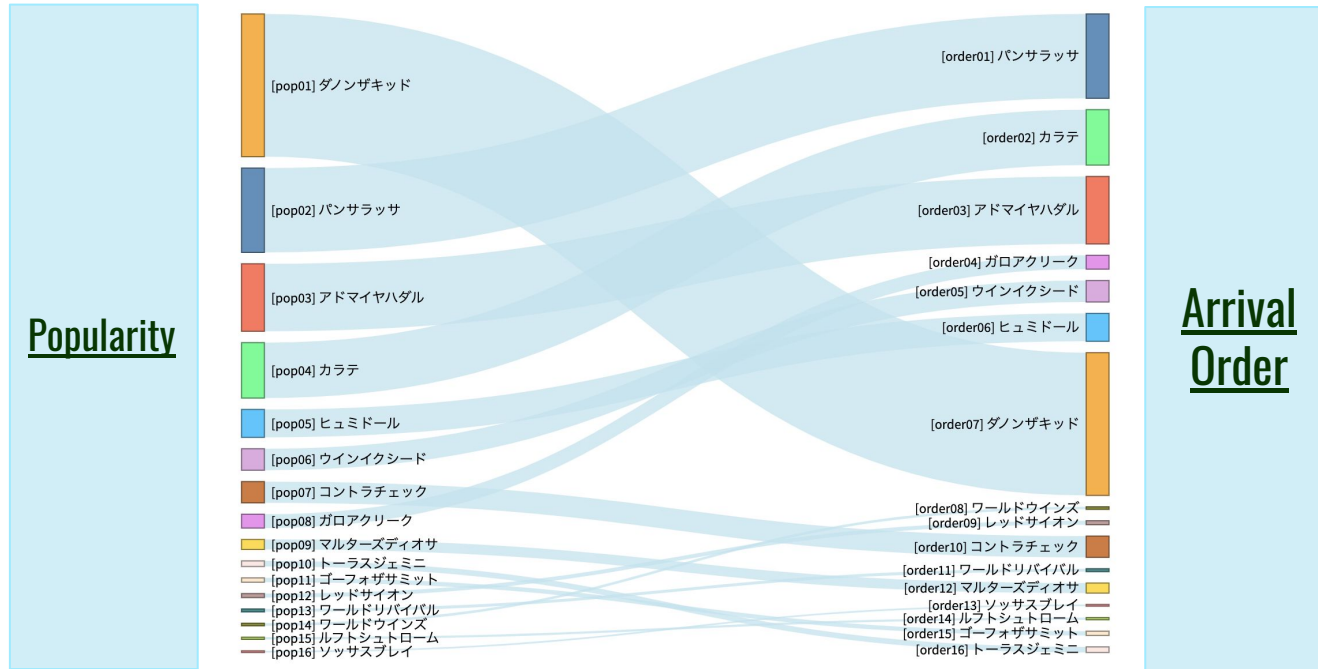
Monitoring Odds (Hourly)



Ex.1) Horse Racing Prediction(3/4): [\[Link\]](#)

Visualize the discrepancy between popularity and arrival order as a back test

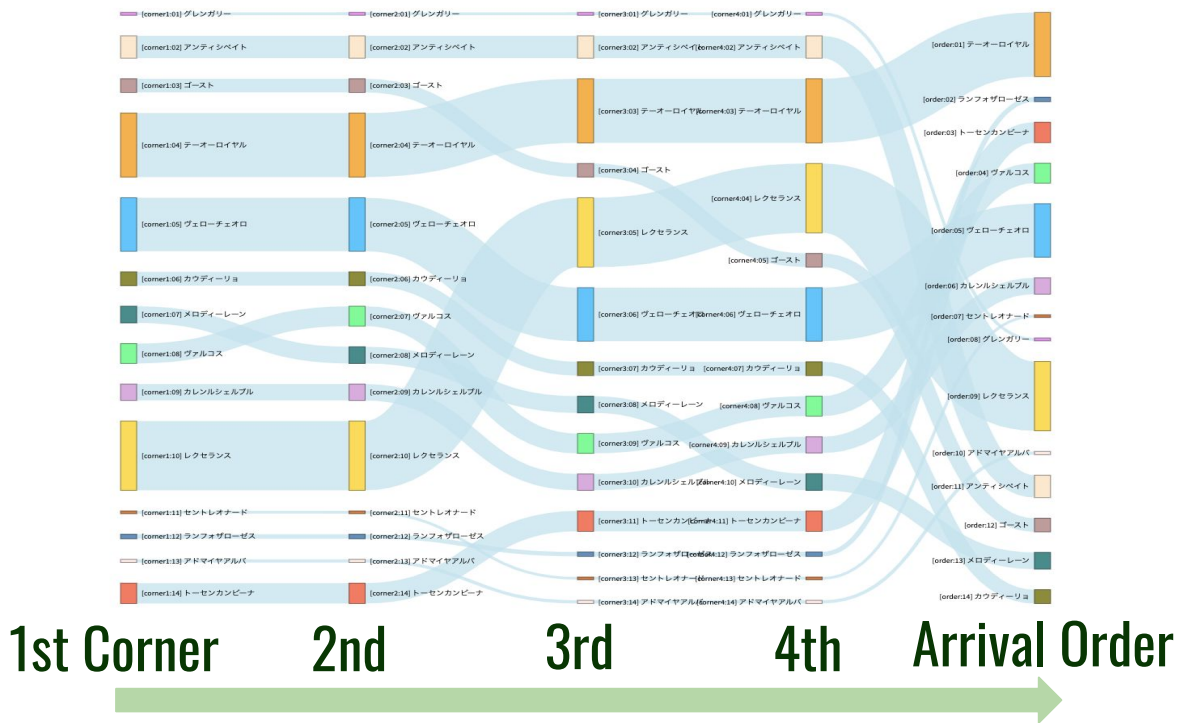
Popularity vs Arrival Order



Ex.1) Horse Racing Prediction(4/4): [\[Link\]](#)

Visualize the order of passing corners, as the race development can be imagined

The Order of Passing Corners



Ex.2) Machine Learning Glossary: [\[Link\]](#)

Create a table report for each category by scraping the glossary

Navigation: [\[Regression Analysis \(3\) \]](#) [\[Response Variable \(3\) \]](#) [\[Sigmoid Function \(3\) \]](#) [\[z-score \(3\) \]](#)

Category: [\[Regression Analysis \(3 terms\) \]](#)

Show entries

Search:

| No | Term | trunc-summary | tag | date |
|----------------------------------|------------------------------------|--|--|----------------------------------|
| <input type="text" value="All"/> | <input type="text" value="All"/> | <input type="text" value="All"/> | <input type="text" value="All"/> | <input type="text" value="All"/> |
| 151 | Standard error | Standard error is the measurement of how dispersed a sample's means are from the population mean. | Regression Analysis / Standard Deviation / Confidence Interval | 05/17/2019 |
| 160 | Heteroscedasticity | Heteroscedasticity refers to data for which the variance of the dependent variable is unequal across the range of independent variables. | Regression Analysis / Variance | 05/17/2019 |
| 192 | F-Distribution | The F distribution is a right-skewed distribution used commonly in another statistical test called an Analysis of Variance (ANOVA). | Probability / Regression Analysis / Variance | 05/17/2019 |

DeepAI | TRY ZENDO^{NEW}

Standard error

Standard Deviation | Statistical Classification

536 | share | edit

What is a Standard Error?

Standard error is the measurement of how dispersed a sample's means are from the population mean. In the vast majority of cases, standard error is defined as the [standard deviation](#) divided by the square root of the sample size.

One exception is in [regression analysis](#), where standard error can refer to both the square root of the reduced chi-squared statistic and the standard error for a regression coefficient, such as [confidence intervals](#).

Ex.3) Leaflet Map: [\[Link\]](#)

Scraping the addresses of medical institutions and put it on the map

【05_小児科 (79件)】

【06_産婦人科 (22件)】

【07_眼科 (35件)】

【08_耳鼻科 (25件)】

【09_歯科 (253件)】

【10_精神科 (20件)】



References



- [R Markdown: The Definitive Guide]
- [Reproducible Reporting with RMarkdown]
- [SportSciData]
- [Smart and Interactive Documents]
- [40 Reports with R Markdown]
- [Tools for Reproducible Research]



End of Document

