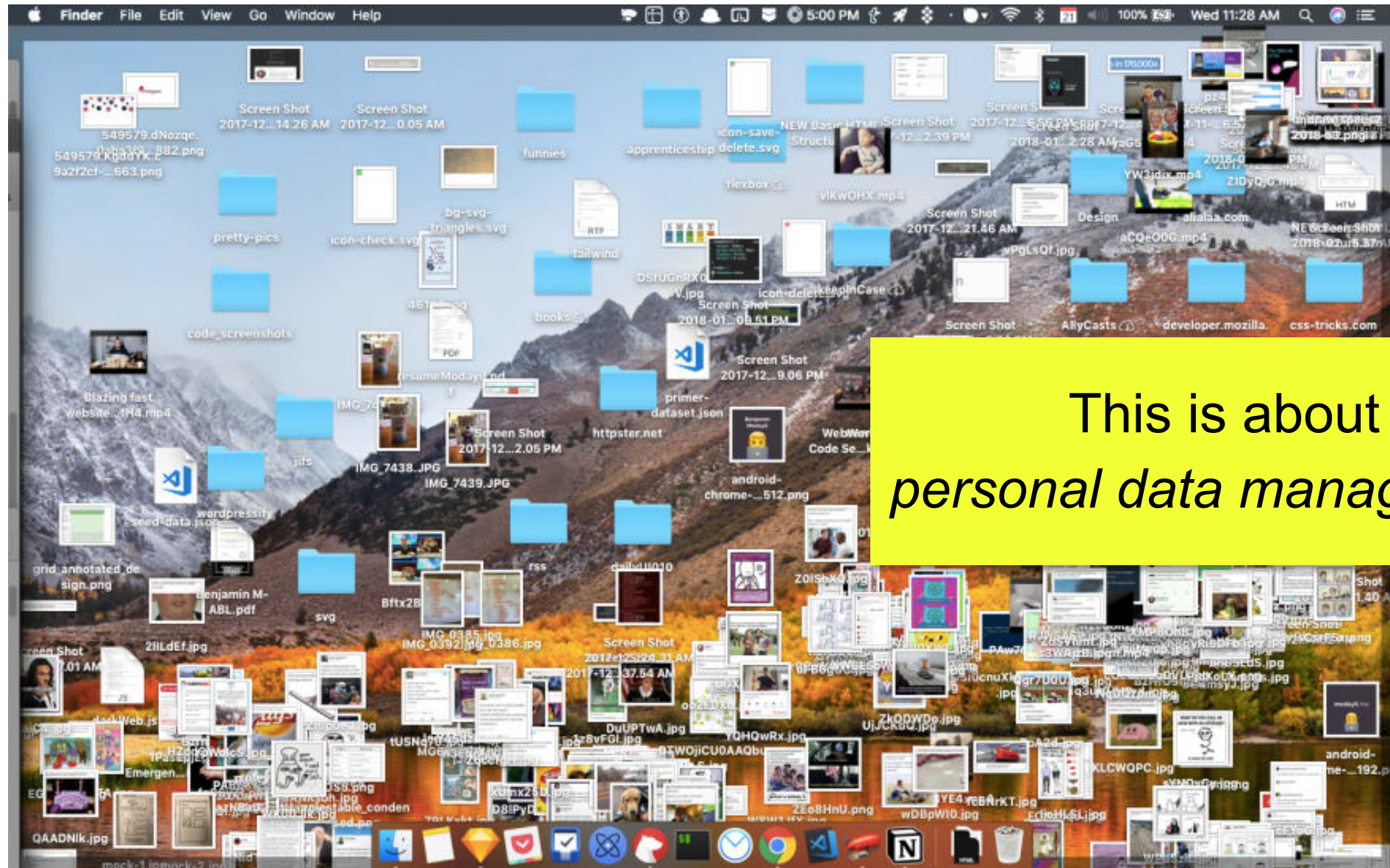


Plain Text Tools

Mirco Schönfeld
University of Bayreuth

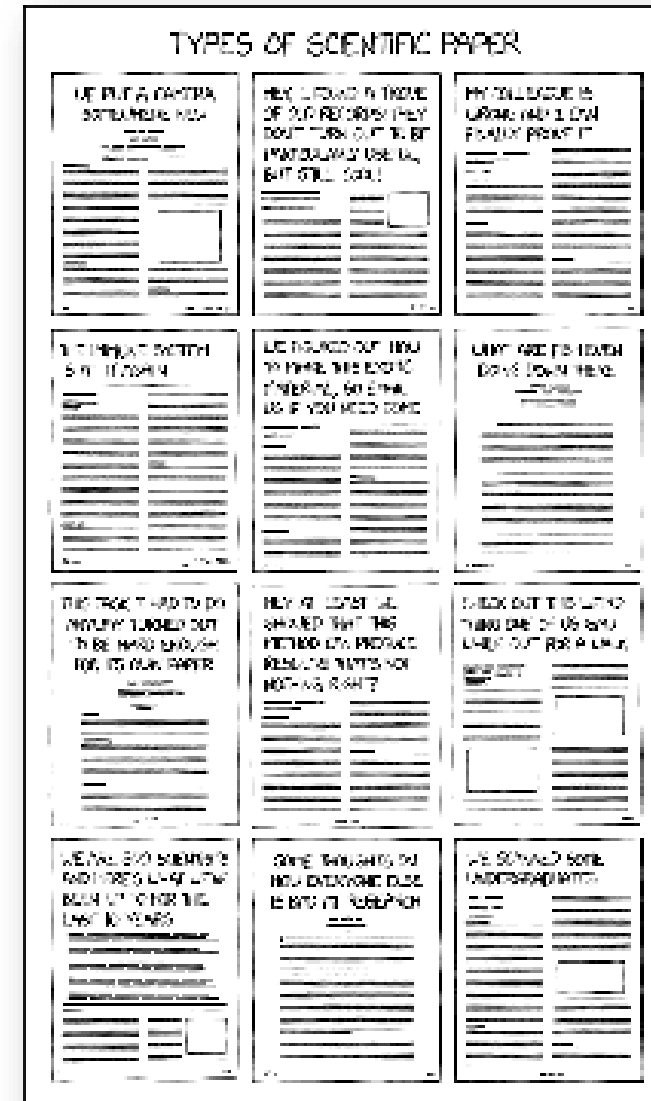
mirco.schoenfeld@uni-bayreuth.de
[@TWlY29](https://twitter.com/TWlY29)



This is about
personal data management

A typical data report

- Cites other papers/journal articles/textbooks
- Presents experimental results/calculations
- Illustrates scenarios using drawings
- References items in the text
- Contains graphs





Basic principles of organizing your work

1. Leave a coherent record of your actions

Sustains reproducibility

Helps organizing references

Your future you will thank you at some point

2. Files and folders need to tell you what they are

Required to organize and document your draft papers, code snippets, datasets, and so on

Also sustains reproducibility

Plain text files are easily searchable – a speaking filename might already be sufficient

3. Automate repetitive and error-prone processes where possible

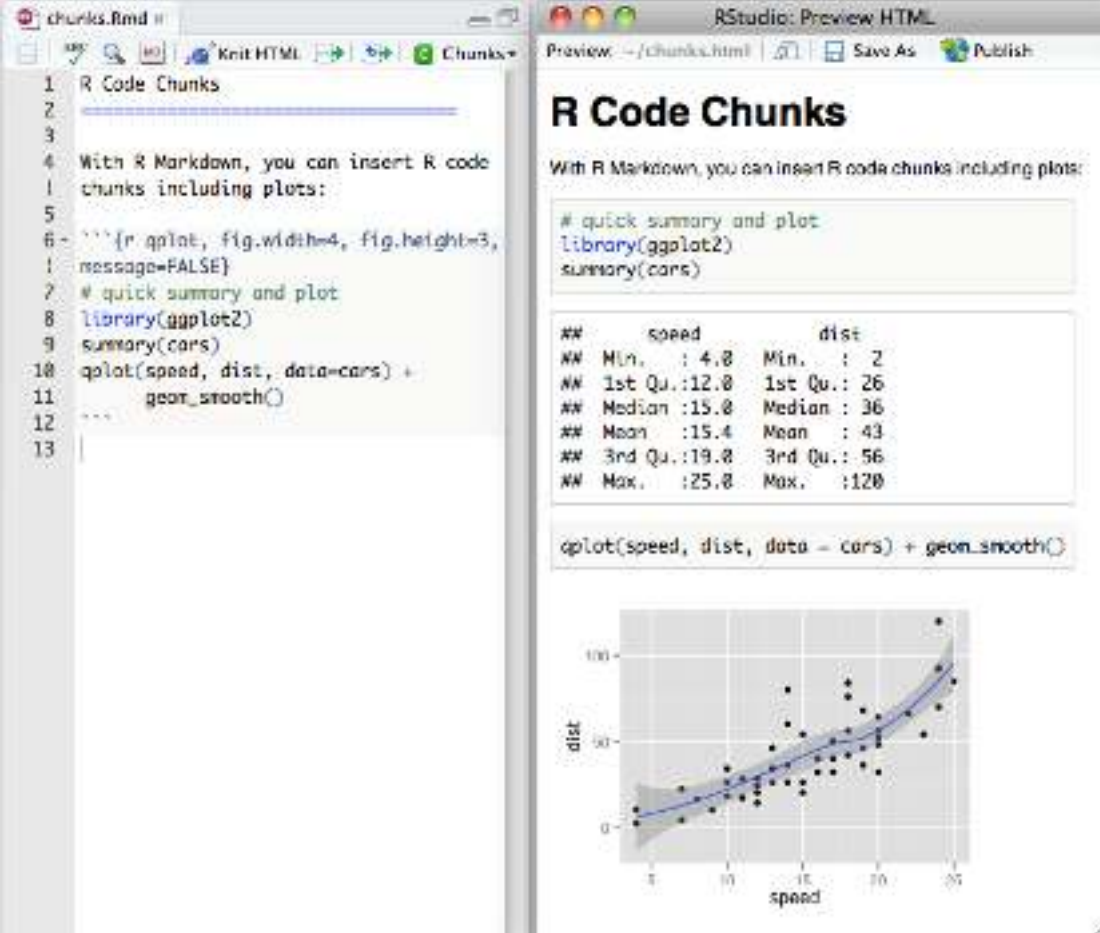
Makes it easier to check for mistakes because there is only one place to correct them

Two models of work

1. Office Model:

Name
Press release for approval.doc
Press release final.doc
Press release FINAL VERSION.doc
Press release FINAL FINAL VERSION.doc
IMPROVED FINAL PRESS RELEASE.doc
REVISED APPROVED FINAL PRESS RELEASE.doc
REVISED APPROVED FINAL PRESS RELEASE v. 2.doc
!! NEW REVISED APPROVED FINAL PRESS RELEASE v. 2.doc
!!! REVISED NEW REVISED APPROVED FINAL PRESS RELEASE v. 2.doc
!!!! Press release as sent.doc

2. Engineering Model:



The screenshot shows the RStudio interface. On the left, the 'chunks.Rmd' editor displays R code chunks. On the right, the 'Preview HTML' window shows the rendered output of these chunks.

chunks.Rmd

```

1 R Code Chunks
2
3
4 With R Markdown, you can insert R code
  chunks including plots:
5
6 ```{r qplot, fig.width=4, fig.height=3,
  message=FALSE}
7 # quick summary and plot
8 library(ggplot2)
9 summary(cars)
10 qplot(speed, dist, data=cars) +
11   geom_smooth()
12
13

```

R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```

# quick summary and plot
library(ggplot2)
summary(cars)

```

##	speed	dist
## Min.	: 4.8	Min. : 2
## 1st Qu.:	:12.8	1st Qu.: 26
## Median :	:15.8	Median : 36
## Mean :	:15.4	Mean : 43
## 3rd Qu.:	:19.8	3rd Qu.: 56
## Max. :	:25.8	Max. :120

```

qplot(speed, dist, data = cars) + geom_smooth()

```

The rendered HTML output includes a scatter plot of distance (dist) versus speed (speed) for the 'cars' dataset. The plot shows a positive correlation between speed and distance, with a blue smoothed trend line and a grey shaded confidence interval.

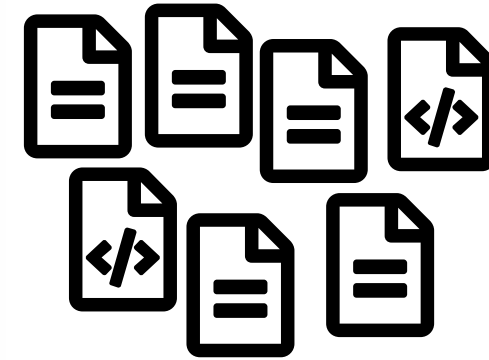
The difference between those models

1. Office Model:



Basic principles of organizing your work are easier to implement using the engineering model.

2. Engineering Model:



A Word file is probably the center of your project

- Changes are tracked *inside* that file
- Citation & reference managers plug into Word
- Output of data analysis is pasted into the file
- Often, the word file *is the output*

The project is organized around a version control repository

- Changes are tracked *outside* of files
- Data analysis is kept in code that produces output in a known manner ideally reproducible
- Citation & references are kept in separate files
- Final output is assembled from a bunch of input files

How to get started

How to draw an owl

1.



1. Draw some circles

2.



2. Draw the rest of the fucking owl

Thanks.

mirco.schoenfeld@uni-bayreuth.de