

# Stat405

Statistical reports

Hadley Wickham

1. (Homework update)
2. More project details
3. Statistical reports

# Project

# Details

4-5 questions. Should be integrated.

10-15 page report

Read the grading rubric!

# Team evaluation

You will grade all members of your team (including yourself) on effort and team citizenship. Each grade must be accompanied by a brief comment.

These will be used to adjust your project grades.

After the project has been handed in, you will have a team meeting to discuss what worked and what didn't.

# High quality comments

He attended group meeting every time, made a serious effort at assigned work and also attempted to make contributions in group meetings.

Missed a meeting and came late another time, often with excuses any of us could provide, since we take the same classes.

She gets her assignments done, but not always by the agreed upon time.

Because we don't have specific responsibility for each person, pretty much all the work is done during group meetings. Each of us just do as much as we can at meetings.

# Useful sites

Dropbox

Google docs

Google groups

# **Statistical reports**

# Statistical reports

Regardless of whether you go into academia or industry, you need to be able to present your findings.

And you should be able to do more than just present them, you should be able to **reproduce** them.

*In one directory*

Data (.csv)

+

Code (.r)

+

Graphics (.png, .pdf)

+

Written report (.tex)

# Working directory

**Terminal** (linux or mac): the working directory is the directory you're in when you start R

**Windows:** File | Change dir.

**Mac:** ⌘-D

**Rstudio:** Tools | Change working dir...

# **Code as communication**

<http://stat405.had.co.nz/r-style.html>

# Code presentation

Use comments (#) to explain your reasoning and to create scannable headings in your code

Every comma should be followed by a space, and every mathematical operator (+, -, =, \*, / etc) should be surrounded by spaces.

Parentheses do not need spaces

Lines should be at most 80 characters. If you have to break up a line, indent the following piece (you can tell it's too long if it doesn't fit on the printed page)

```
qplot(table, depth, data=diamonds)
qplot(table, depth, data=diamonds)
+xlim(50, 70)+ylim(50, 70)
qplot(table~depth, data=diamonds, geom="histo
gram")
qplot(table/depth, data=diamonds, geom="histo
gram", binwidth=0.01)+xlim(0.8, 1.2)
```

```
# Table and depth -----
```

```
qplot(table, depth, data = diamonds)  
qplot(table, depth, data = diamonds) +  
  xlim(50, 70) + ylim(50, 70)
```

```
# Is there a linear relationship?
```

```
qplot(table - depth, data = diamonds,  
  geom = "histogram")
```

```
# This bin width seems the most revealing
```

```
qplot(table / depth, data = diamonds,  
  geom = "histogram", binwidth = 0.01) +  
  xlim(0.8, 1.2)
```

```
# Also tried: 0.05, 0.005, 0.002
```

```
# Table and depth -----
```

```
qplot(table, depth, data = diamonds)  
qplot(table, depth, data = diamonds) +  
  xlim(50, 70) + ylim(50, 70)
```

```
# Is there a linear relationship?
```

```
qplot(table - depth, data = diamonds,  
  geom = "histogram")
```

```
# This bin width seems the most revealing
```

```
qplot(table / depth, data = diamonds,  
  geom = "histogram", binwidth = 0.01) +  
  xlim(0.8, 1.2)
```

```
# Also tried: 0.05, 0.005, 0.002
```

# Your turn

Open the style drills at <http://stat405.had.co.nz/drills/style.html>

Rewrite the first example to comply with the style recommendations.

# **Written report**

# Latex

We are going to use the open source document typesetting system called **latex** to produce our reports. From now on, all projects and homeworks **must** be done in latex.

Latex widespread in statistics – if you ever write a journal article, you will probably write it in latex.

# Edit-Compile-Preview

Edit: a text document with special formatting

Compile: to produce a pdf

Preview: with a pdf viewer

One click in RStudio. If you're not using RStudio, you'll need to figure out how to do this yourself.

# Latex

Template

Sections

Images

Figures and cross-references

Verbatim input (for code)

~/Desktop/sample-report - RStudio

template.tex x

Format Compile PDF

```
2 \documentclass[oneside]{article}
3 % Decrease the margins a little
4 \usepackage{fullpage}
5
6 % Set up for including graphics
7 % We'll use png or pdf graphics
8 \usepackage[pdftex]{graphicx}
9 \DeclareGraphicsExtensions{.png,.pdf}
10
11 % Hyperref adds hyperlinks to the document automatically
12 % It's not much use yet, but it will be
13 \usepackage{hyperref}
14
15 % For including code into the document
```

17:1 TeX

Console ~/Desktop/sample-report/

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

> |

~/Desktop/sample-report - RStudio

template.tex ×

Format Compile PDF

```
| template and modifying it to meet your needs. As the semester goes on
| you'll learn more about exactly what you're doing and explore more of
| the many advanced features of latex.
34
35 | As you can see from this text, basic text in latex is very easy. You
| just type normal paragraphs (with a blank line between them). Unlike
| word, you don't need to worry about the appearance of the text - latex
| takes care of laying it out for you. For special formatting, use {\bf
| bold} or {\it italics} or \verb|some_R_code()|. You'll notice that in
| latex most special formatting commands use the backslash and brackets -
| \verb|\\, \verb|{| and \verb|}|.
36
37 | Quotes are a little tricky: use `` '' (\verb|Shift + ~| on the left), or
| ` ', but never "".
```

35:1 TeX

Console Compile PDF ×

.../sample-report/template.tex View Log

Output Issues

- ✘ Line 35 Undefined control sequence.
- ✘ Line 35 Missing \$ inserted.
- ✘ Line 35 Double subscript.
- ✘ Line 36 Missing \$ inserted.
- ✘ Line 35 Overfull \hbox (280.55917pt too wide) in paragraph at lines 35--36
- ✘ Line 66 Overfull \hbox (14.99998pt too wide) in paragraph at lines 66--68
- ✘ Line 78 Overfull \hbox (2.10999pt too wide) in paragraph at lines 78--80

# Common errors

Undefined control sequence = you've mistyped a latex command

Missing \$ inserted = you've written something that latex thinks is math (i.e. `_`).  
Convert to `\_` or put inside `$$` or `\verb| |`

`\begin{something}` ended by  
`\end{something else}` = you've forgotten  
`\end{something}`

# Your turn

Download the sample report and unzip it.  
Open in Rstudio and run it.

Start on homework 3.