Workflow

Hadley Wickham

October 2009

In one directory

=

Written report

+

Code

+

Data

Working directory

Remember to set your working directory.

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

On windows: `setwd(choose.dir())`

On the mac: ⌘-D

Saving data

```
# For long-term
write.table(slots, file = "slots-3.csv",
  sep=",", row = F)

# For short-term caching
save(slots, file = "slots.rdata")
```

.csv	.rdata
`read.csv()`	`load()`
`write.table(sep = ",", row = F)`	`save()`
Only data frames	Any R object
Can be read by any program	Only by R
Long term	Short term caching of expensive computations

Code is

communication!

```
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
```

```r
# 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
```

Aside: strategy

The biggest problem I see new programmers make is trying to do too much at once.

Break the problem into pieces and solve the smallest piece first. Then check each piece before solving the next problem.