LANCASTER UNIVERSITY

Working Habits

How to make your life easier and more efficient.

b.rowlingson@gmail.com
@geospacedman

Faculty of Health and Medicine,
Lancaster University

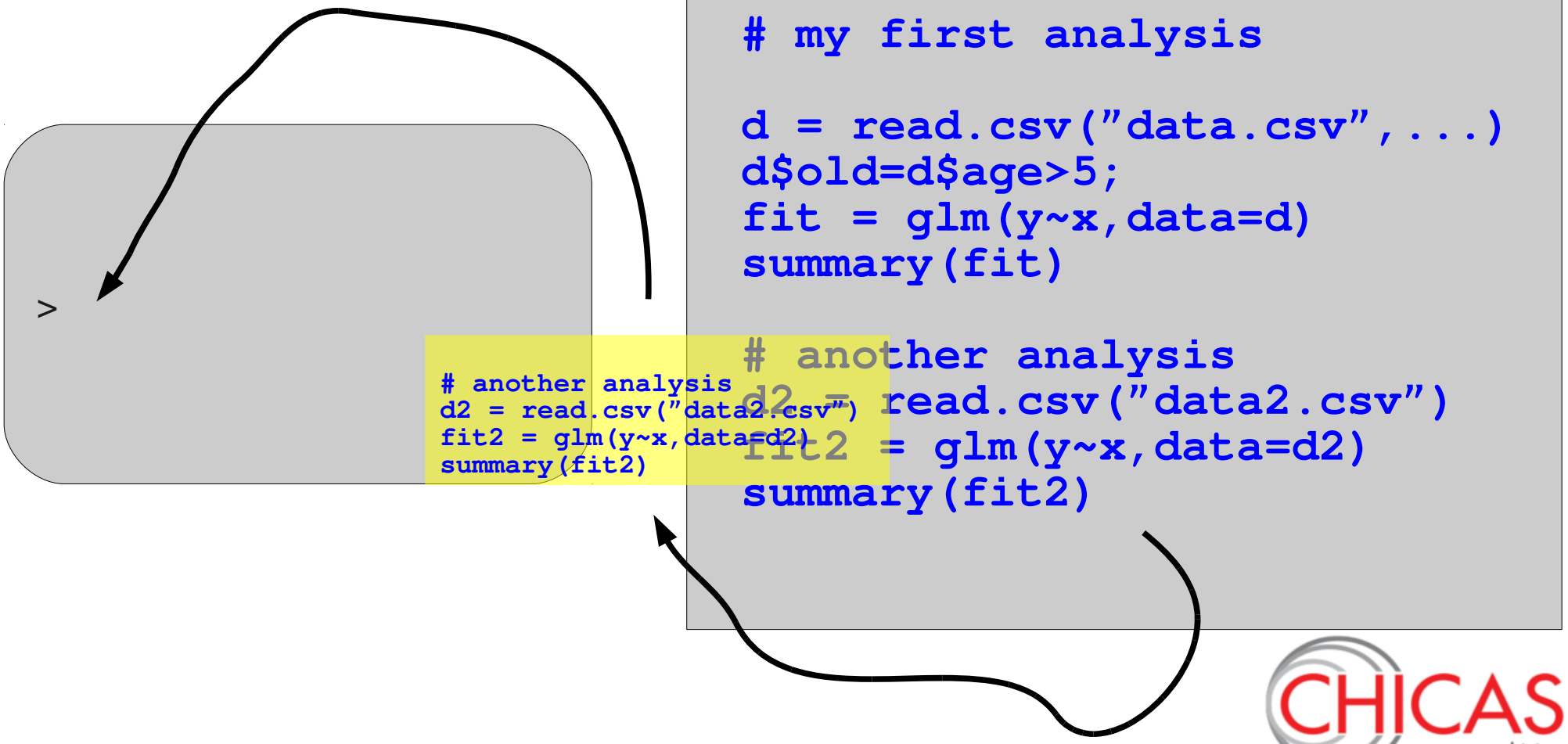
CHICAS
combining health information, computation and statistics

Command Line

```
> readmydata=function(f,...){
    d = read.csv(f,...)
    d$old=d$age>5;
    d
  }
> data = readmydata("data.csv")
> fit = glm(y~x,data=data)
> summary(fit)
```

CHICAS
combining health information, computation and statistics

Cut and Paste



CHICAS
combining health information, computation and statistics

Source

first.R

```
# my first analysis
d = read.csv("data.csv",...)
d$old=d$age>5;
fit = glm(y~x,data=d)
summary(fit)
```

```
>source("first.R")
```

second.R

```
# another analysis
d2 = read.csv("data2.csv")
fit2 = glm(y~x,data=d2)
summary(fit2)
```

CHICAS combining health information, computation and statistics

(One) function per file

```
> source('first.R')
> first('data.csv')
> source('second.R')
> second('data.csv')
> second('data2.csv')
```

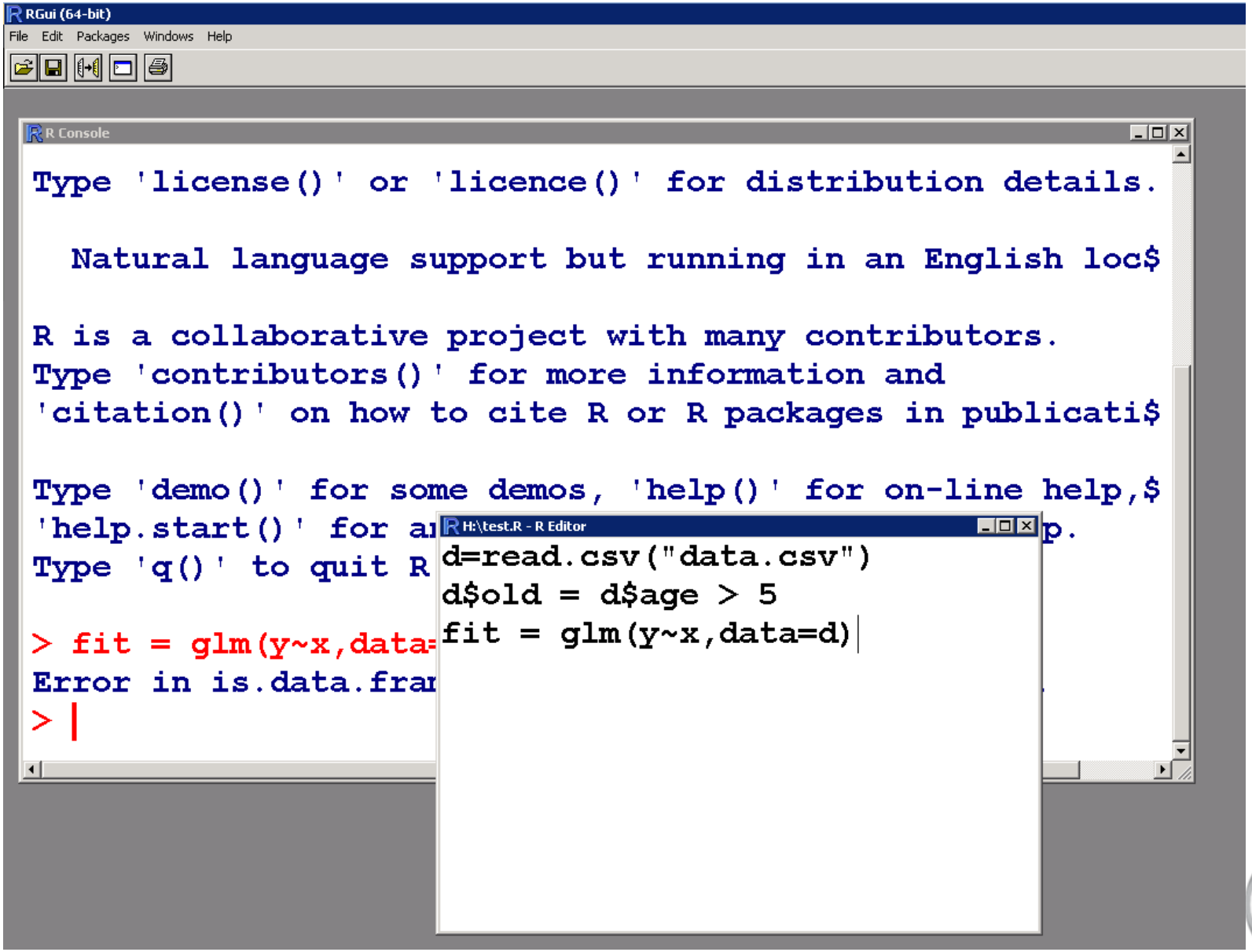
first.R

```
# my first analysis
first=function(f){
d = read.csv(f,...)
d$old=d$age>5;
fit = glm(y~x,data=d)
summary(fit)
}
```

second.R

```
# another analysis
second=function(f){
d2 = read.csv(f)
fit2 = glm(y~x,data=d2)
summary(fit2)
}
```

Windows GUI

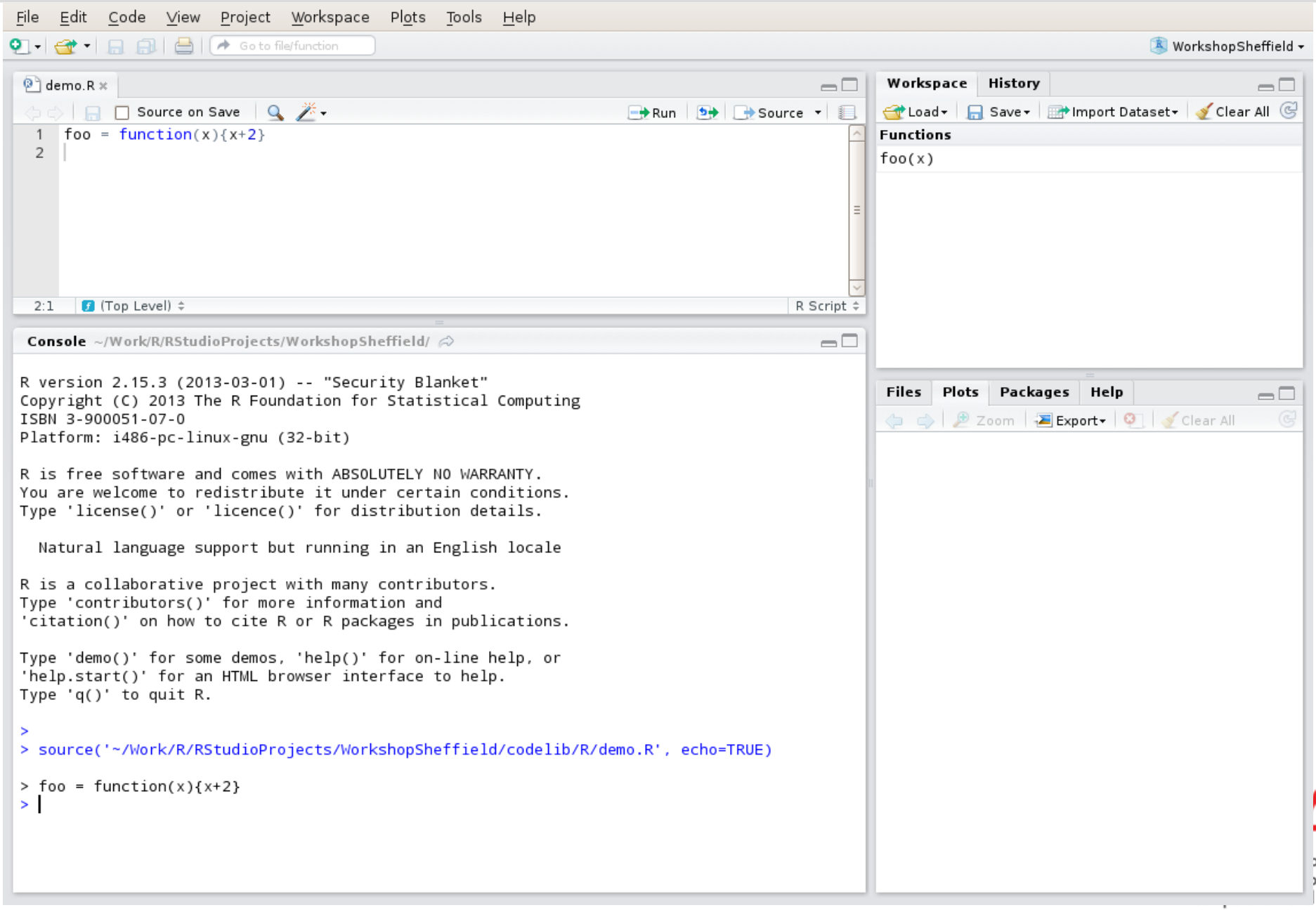


LANCASTER UNIVERSITY

CHICAS
combining health information, computation and statistics

RStudio

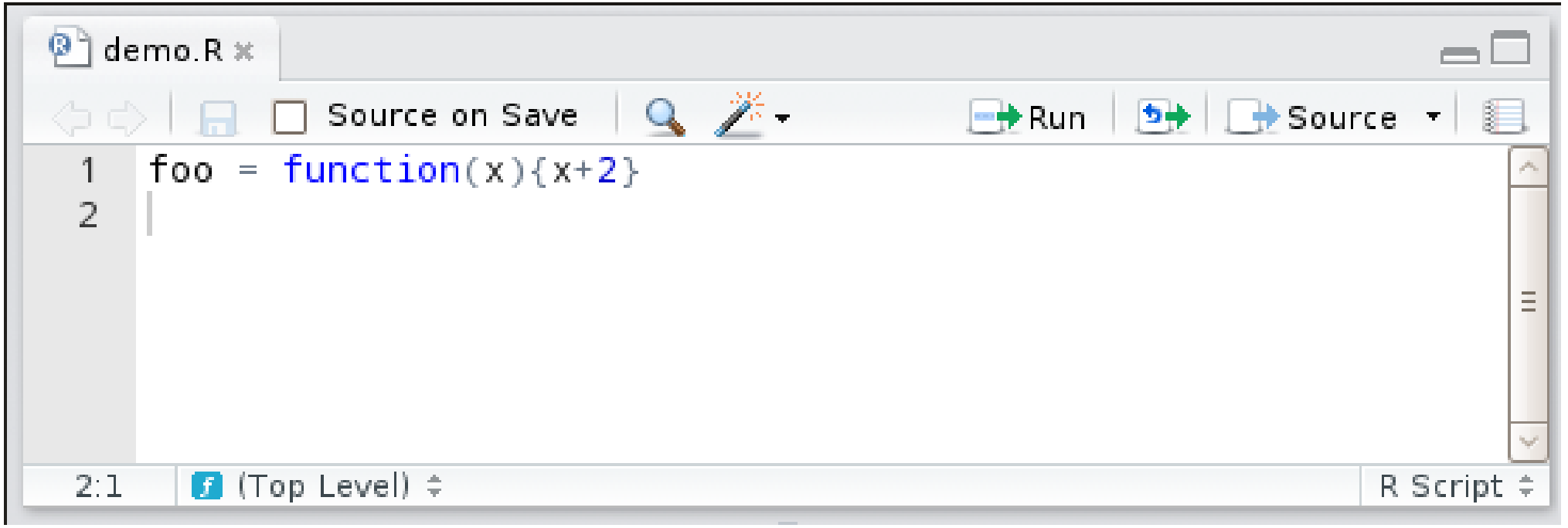
LANCASTER UNIVERSITY



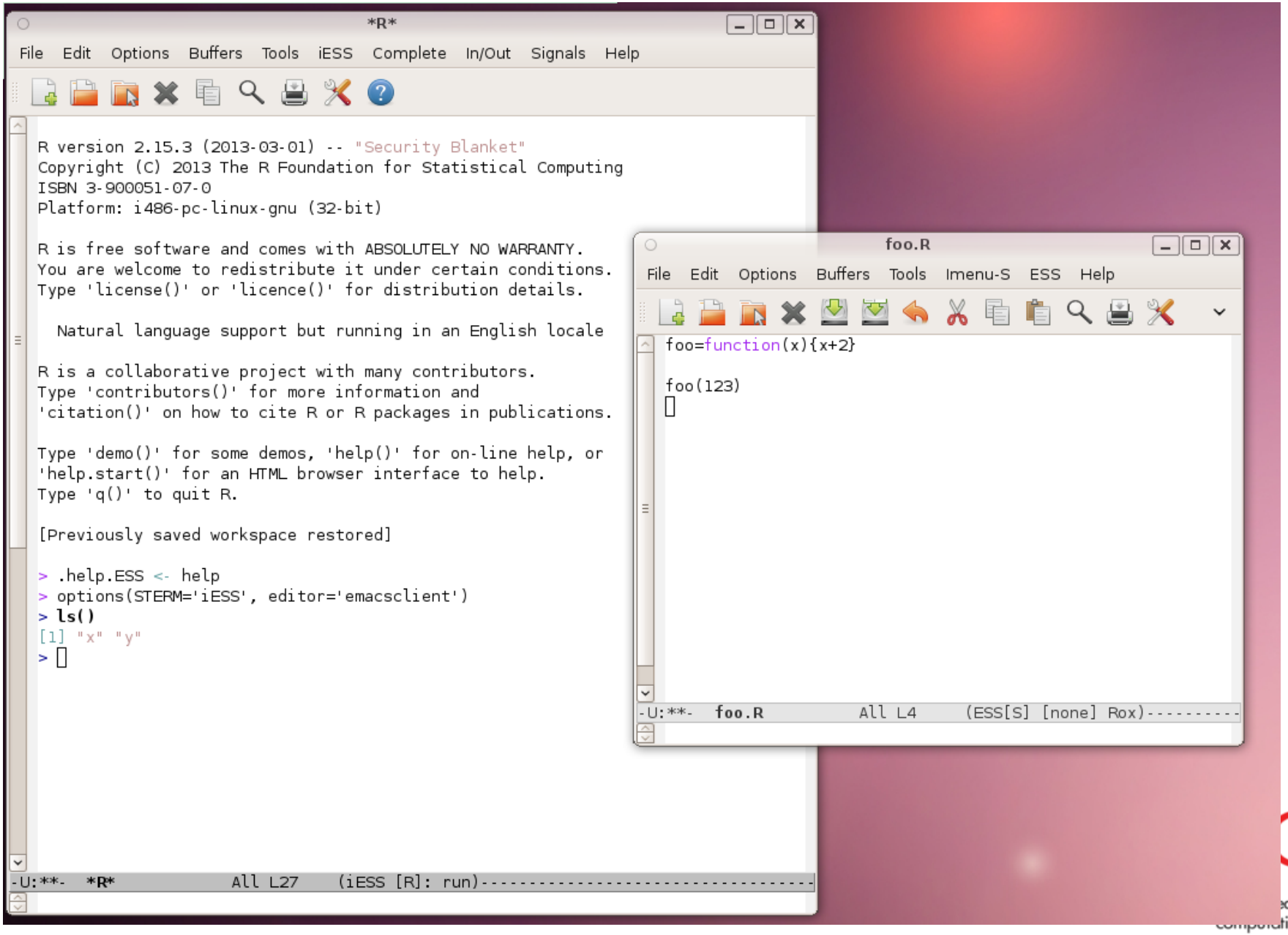
AS

ombining
rmation,
statistics

RStudio


demo.R
Source on Save
Run
Source
1 foo = function(x){x+2}
2
2:1
(Top Level)
R Script

Emacs/ESS



Source problem

LANCASTER UNIVERSITY

- Everything is in **`ls()`**
- Need to **`save`** everything
- Or **`source`** everything
- Stuff copied everywhere
- Solution: ***organise things nicely***

CHICAS
combining health information, computation and statistics

Package devtools

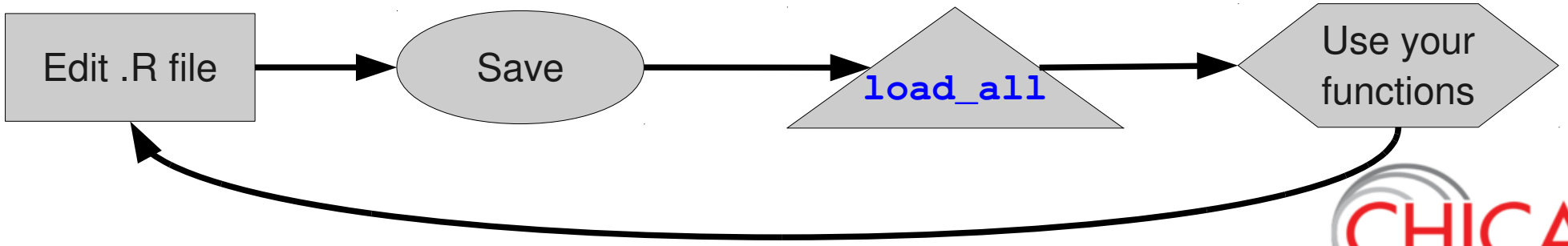
Be functional.

Write functions not scripts

```
> require(devtools)
> create('Code/mycode')
```

Now edit files in **Code/mycode/R/whatever.R**

```
> load_all('Code/mycode')
```



CHICAS
combining health information, computation and statistics

Why bother?

LANCASTER UNIVERSITY

```
> load_all("Code/mycode")
Loading mycode
Adding files missing in collate: test.R

> z = foo(1,2,3)

> ls()
[1] 'z'

> search()
 [1] ".GlobalEnv"        "package:mycode"    "package:roxygen2"
 [4] "package:digest"    "package:devtools"  "package:stats"
 [7] "package:graph"     "package:grDevices" "package:utils"
[10] "package:data"      "package:methods"   "Autoloads"
[13] "package:base"

> ls(pos=2)
[1] "foo"
```

Load my code

Here's my new function

Where's foo?

There it is

CHICAS combining health information, computation and statistics

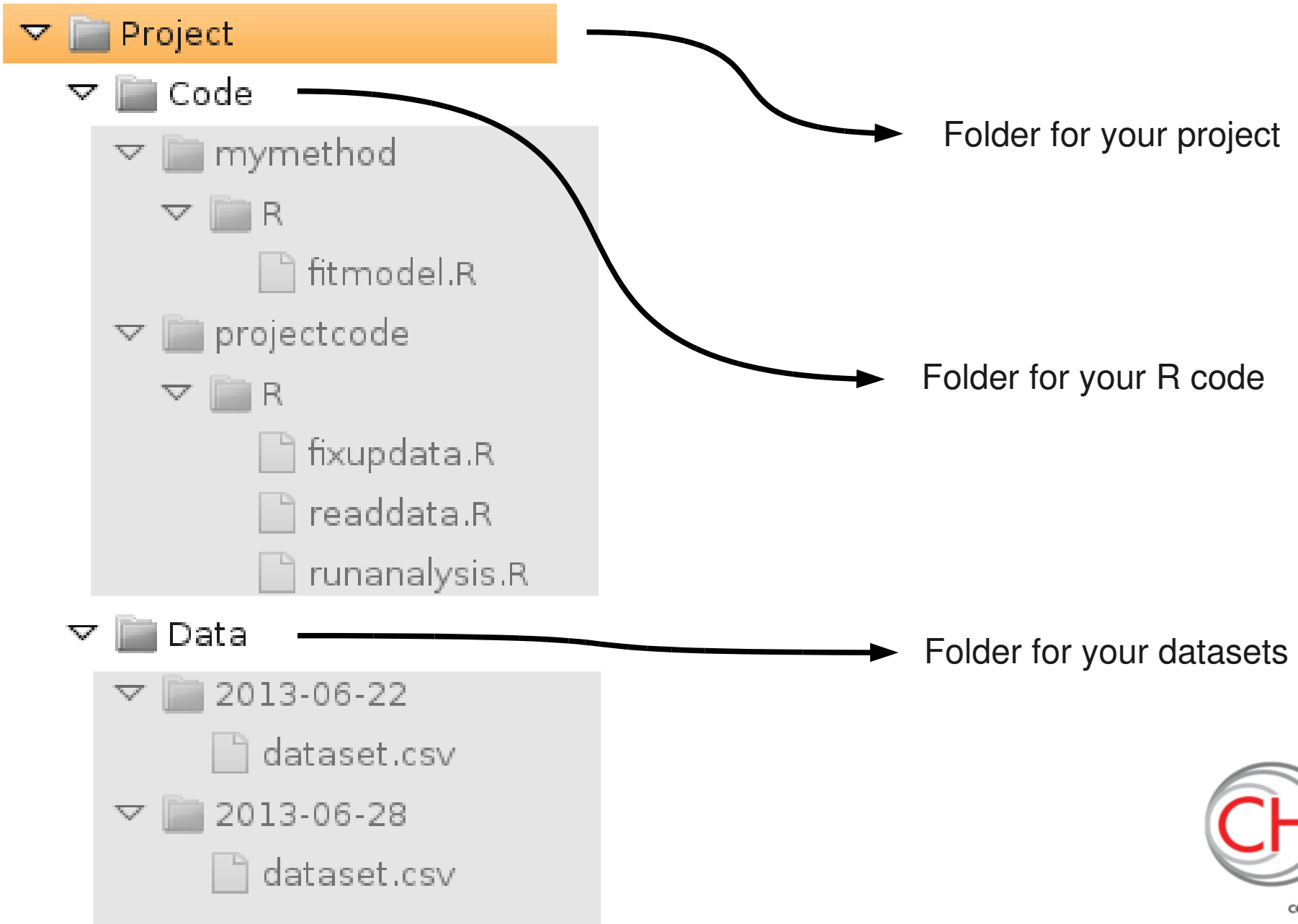
Package devtools

LANCASTER UNIVERSITY

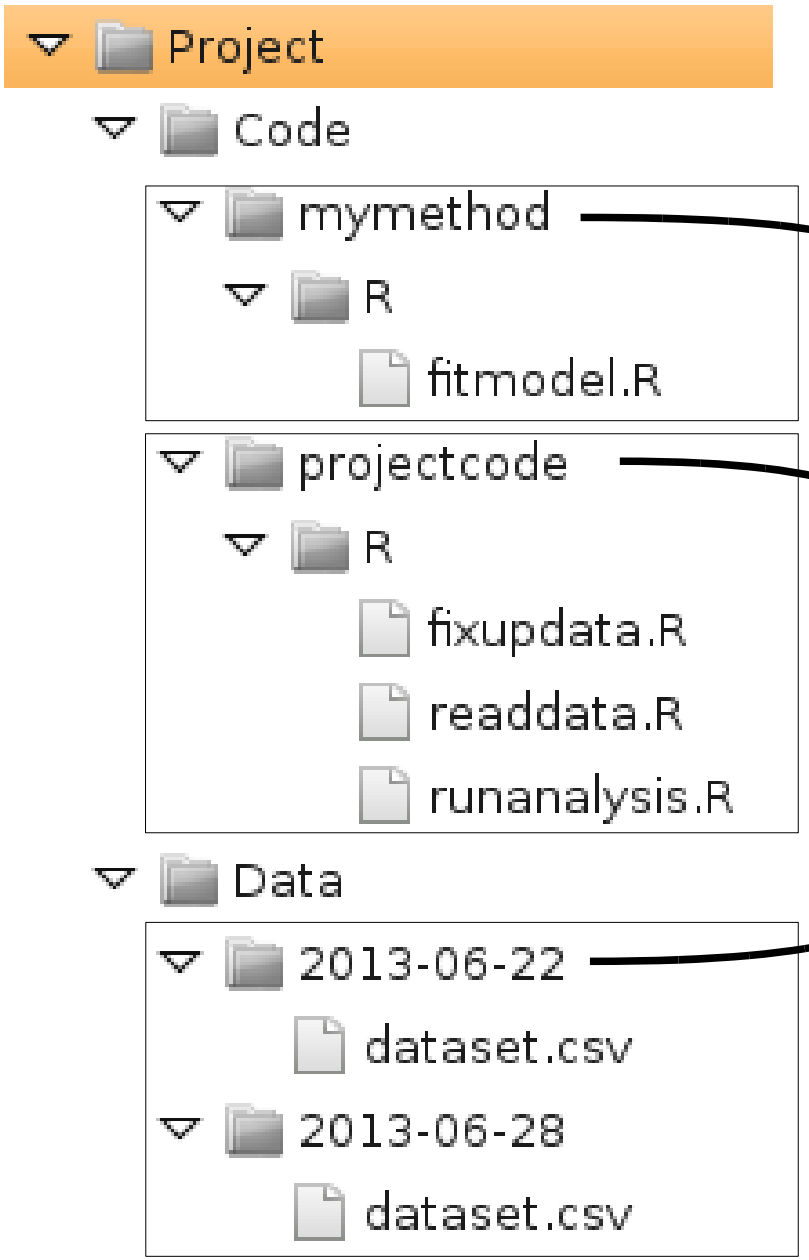
You've just written an R package

CHICAS
combining
health information,
computation and statistics

Handy conventions



Handy conventions



Code that might have wide use

Code only useful in this project

Create a new folder when you get updated datasets. Keep 'readme' files

LANCASTER UNIVERSITY

software carpentry

CHICAS combining health information, computation and statistics

Spin your scripts

LANCASTER UNIVERSITY

```
#'
#' This is my project
#'
#' First I create some data

#+ create
d = data.frame(x=runif(10),y=runif(10))

#' now we do something else

#+ testsum
sum(d$x) # this is a number

#+
plot(d$x,d$y)

#' we can now look at this plot
```

```
> source('makeplot.R')
> require(knitr); spin('makeplot.R')
```

CHICAS
combining health information, computation and statistics

Gmail | R Output | Home Page

file:///home/me/output/project.html

Search

This is my project

First I create some data

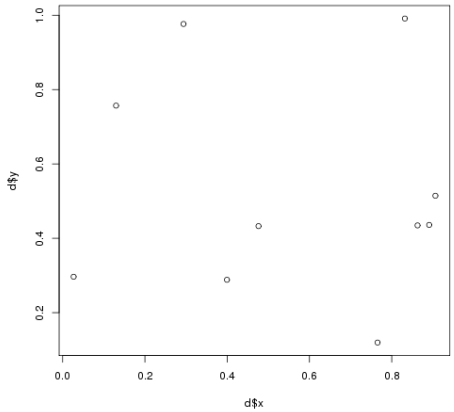
```
d = data.frame(x = runif(10), y = runif(10))
```

now we do something else

```
sum(d$x)  # this is a number
```

```
## [1] 5.583
```

```
plot(d$x, d$y)
```



we can now look at this plot

The `knitr` package

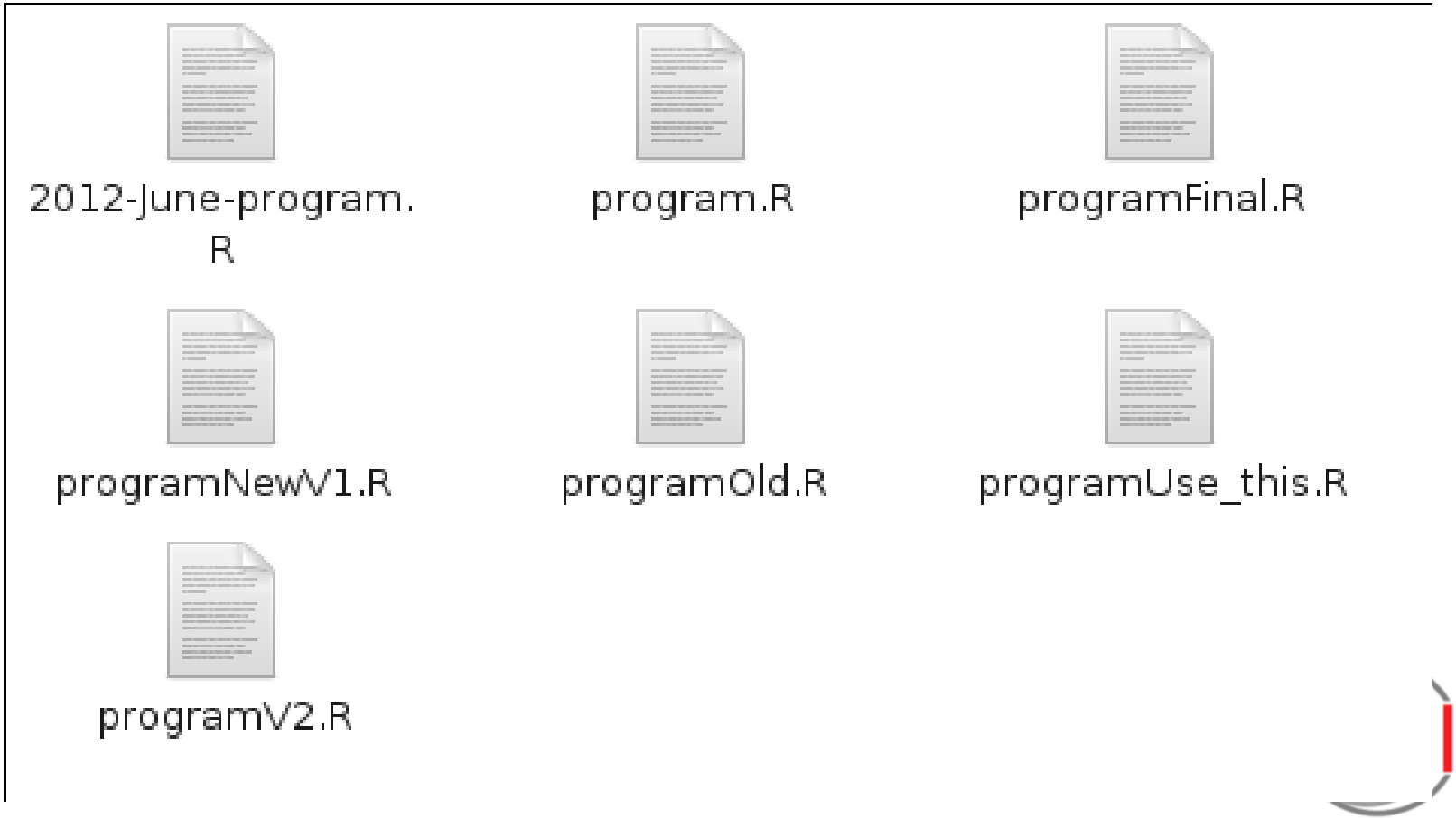
LANCASTER UNIVERSITY

- You put R code in a document
- Knitr processes the document
- Knitr embeds results and graphics in the output
- Can produce HTML, LaTeX, PDF, Markdown, OpenOffice, MS Word
- Works with other languages
- Great for routine reporting, data-driven documents
- Reproducible Research

CHICAS
combining health information, computation and statistics

Revision control

- Disk space is now cheap
- Why delete when you can keep?



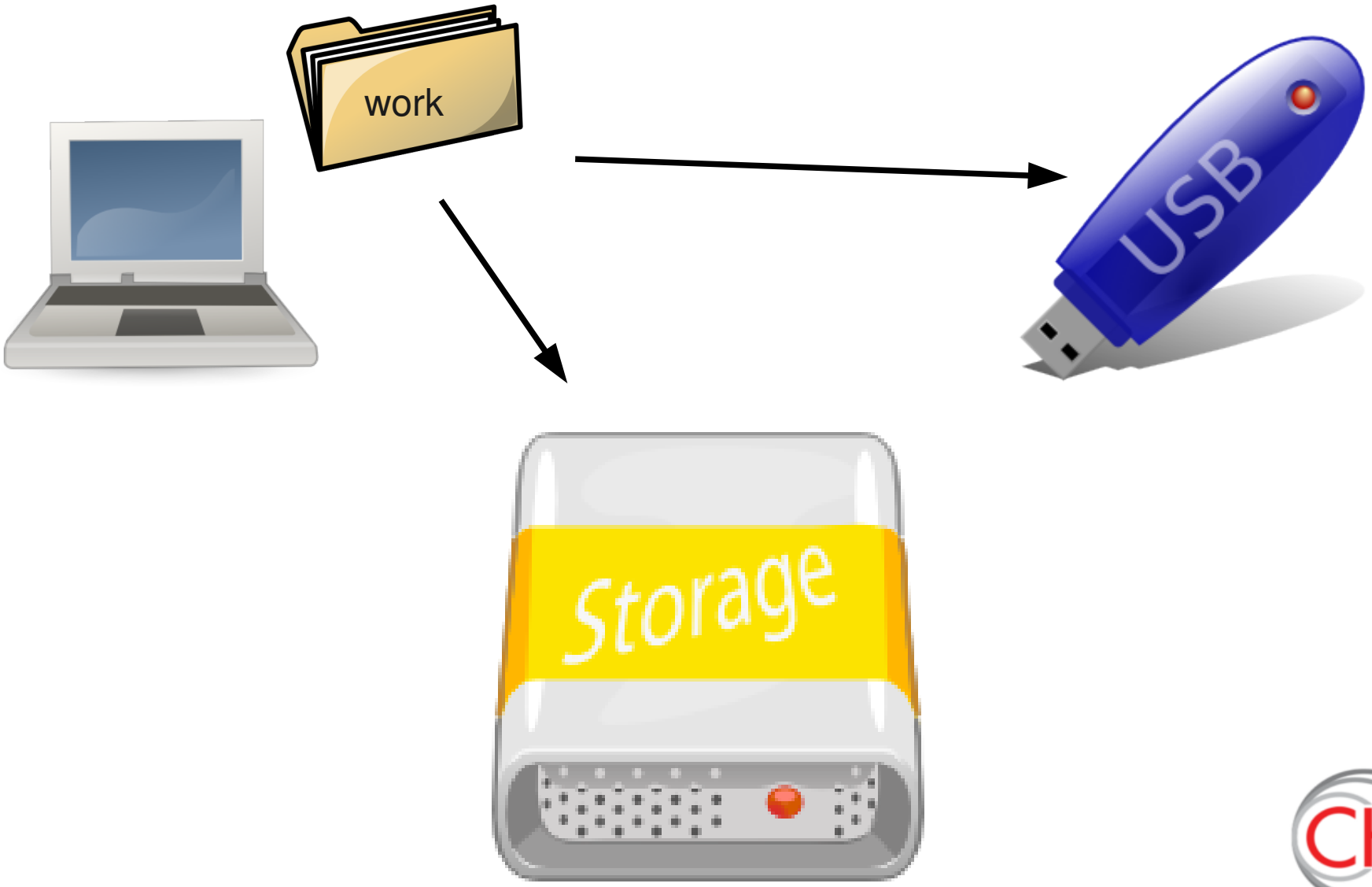
What you really want...

- What did this file look like
  - Last Week?
  - Last Year?
  - When I did the analysis for the paper?
- What's changed today?
- I want to try something, but it might not work out...
  - How can I go back?

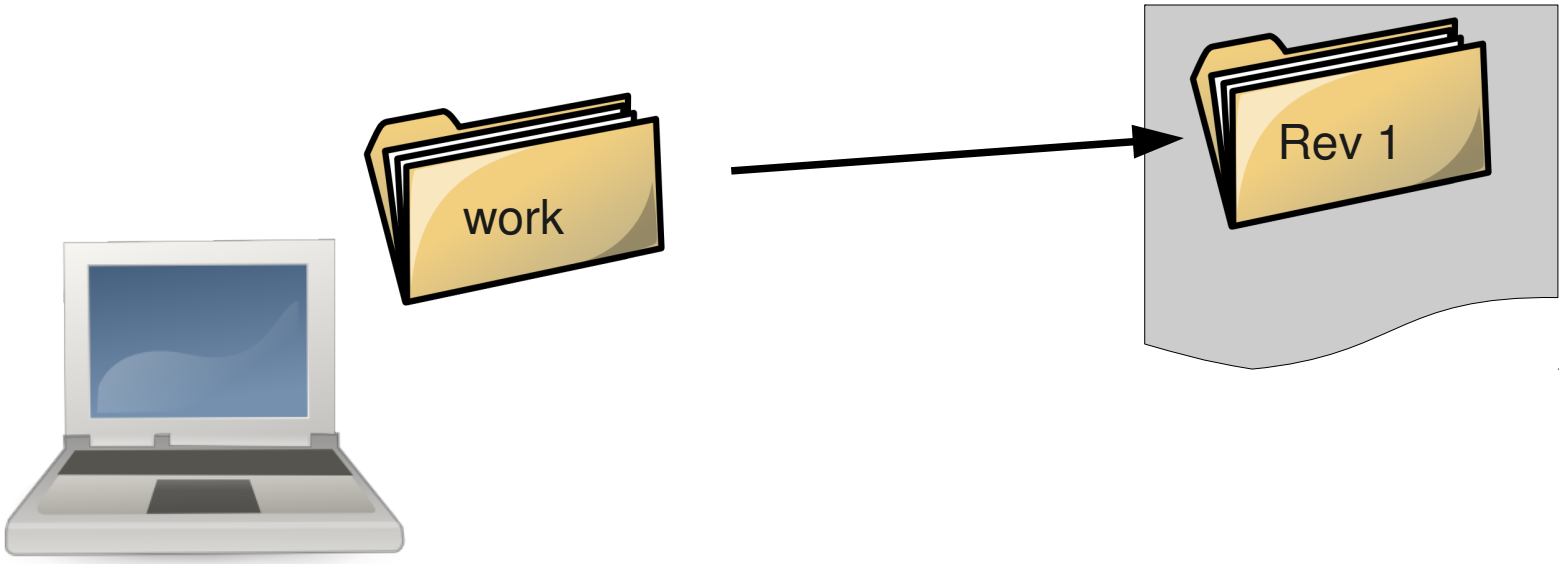
LANCASTER UNIVERSITY

CHICAS
combining health information, computation and statistics

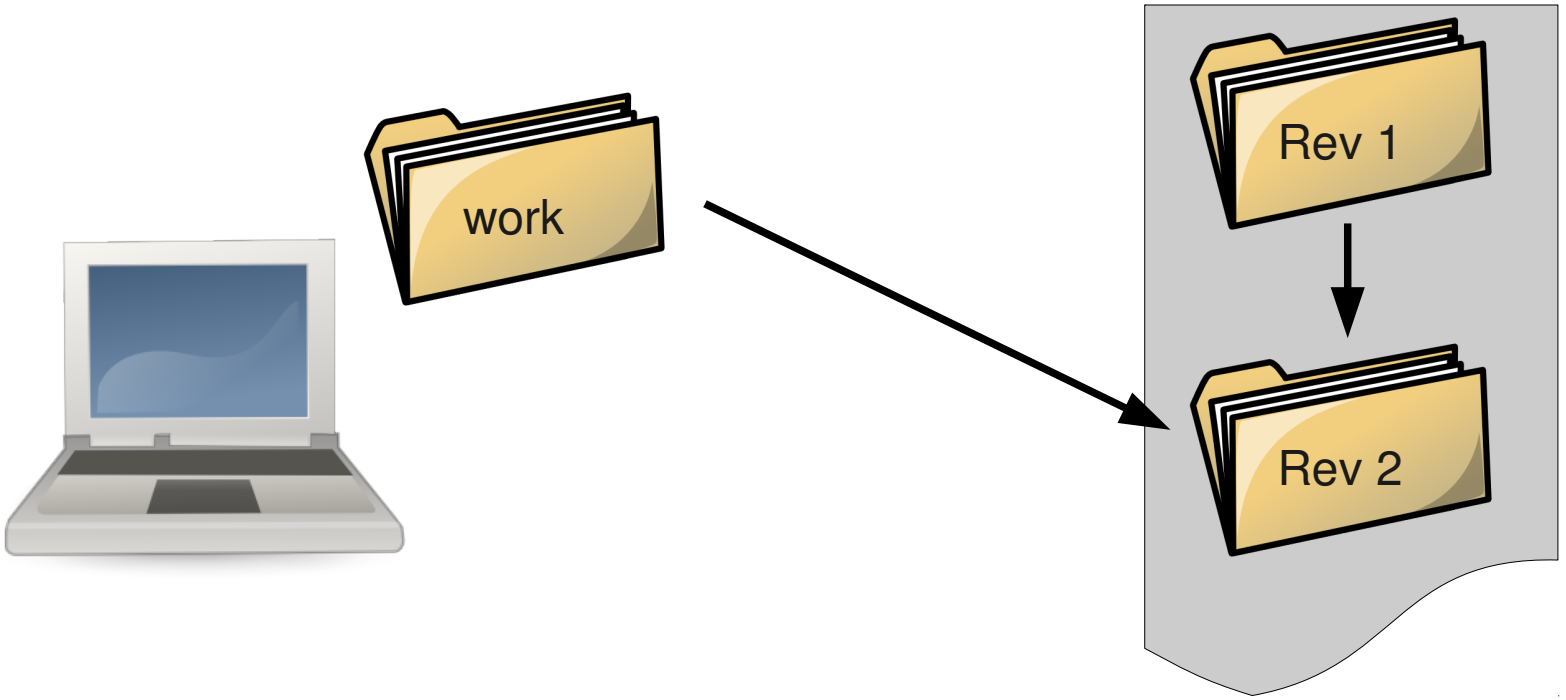
Is This Just Backup?



initialise



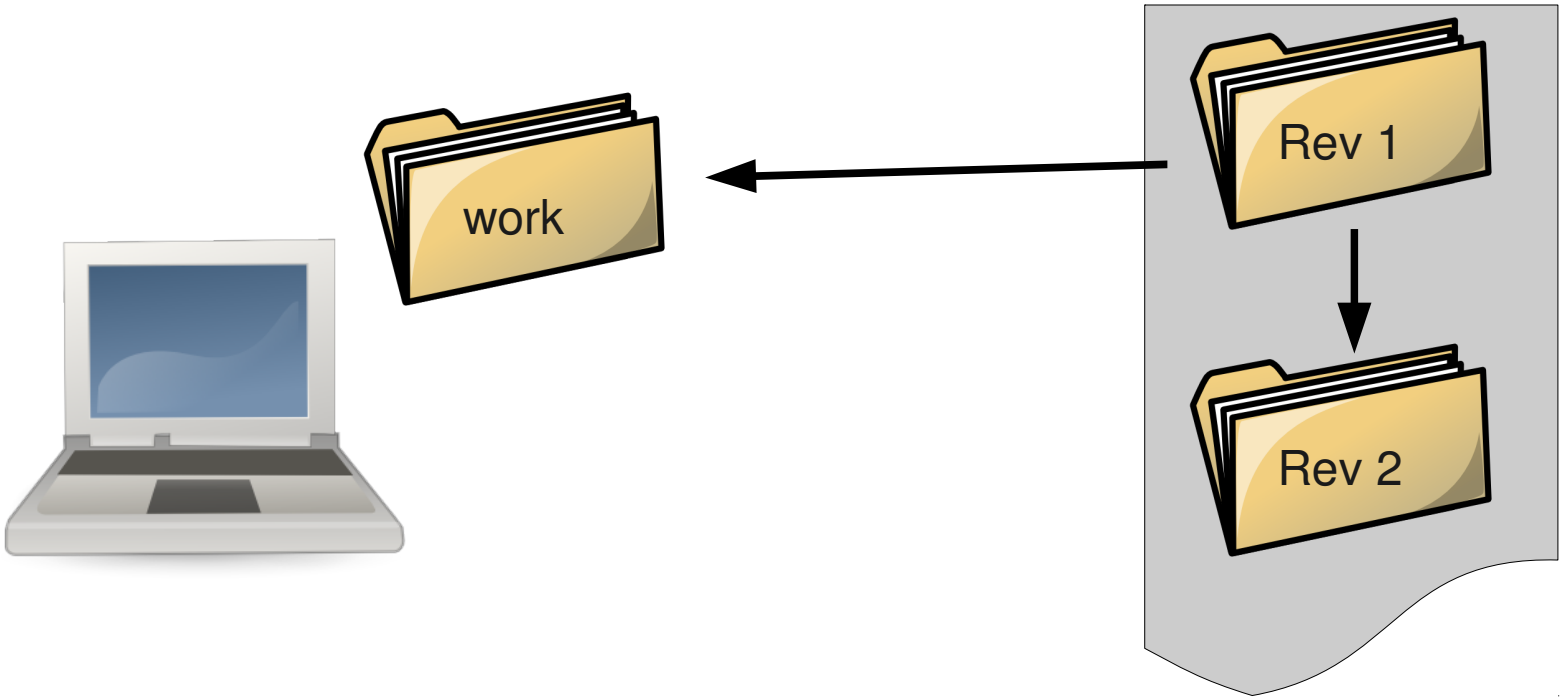
commit



LANCASTER UNIVERSITY

CHICAS
combining
health information,
computation and statistics

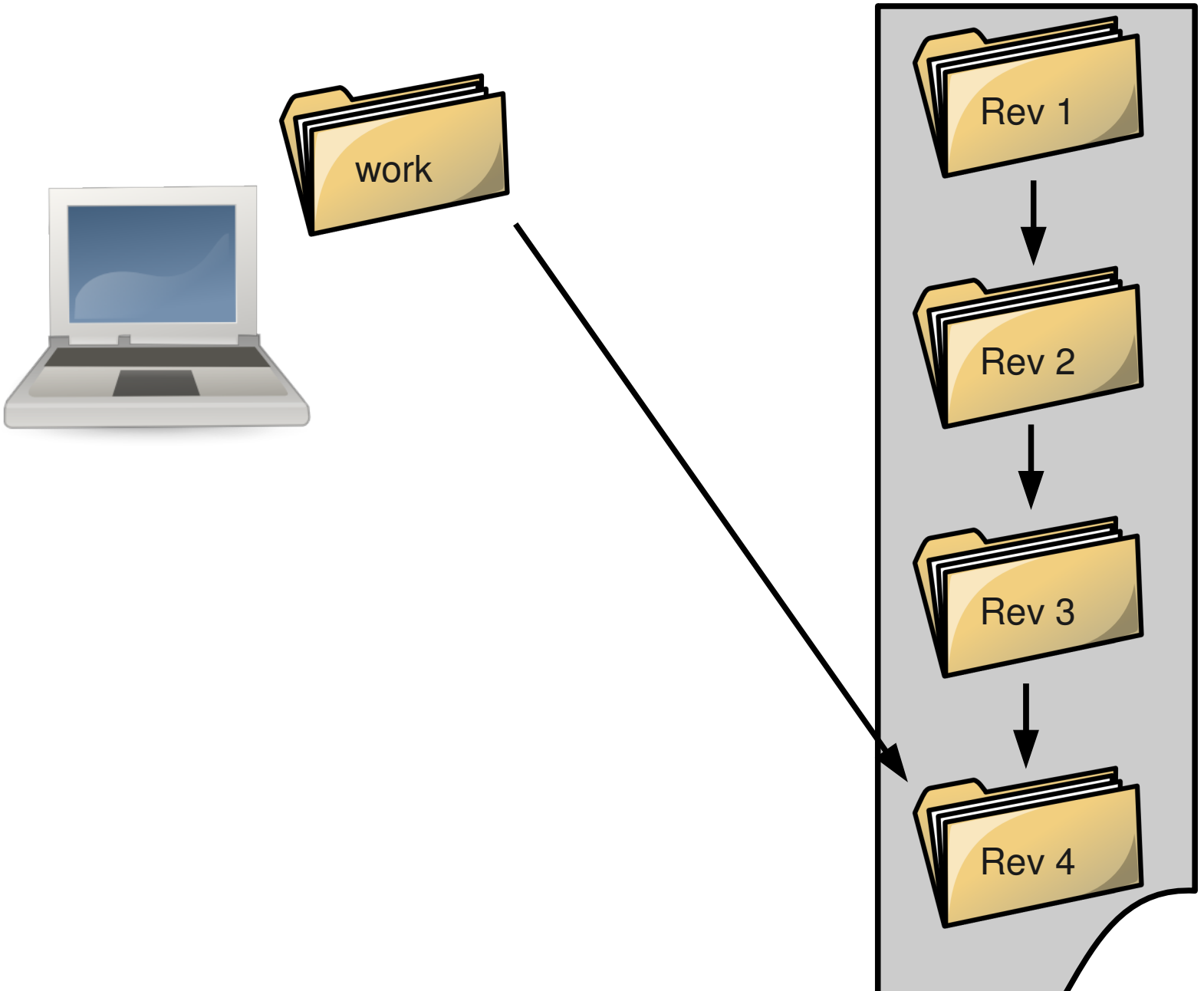
pull



LANCASTER UNIVERSITY

CHICAS
combining
health information,
computation and statistics

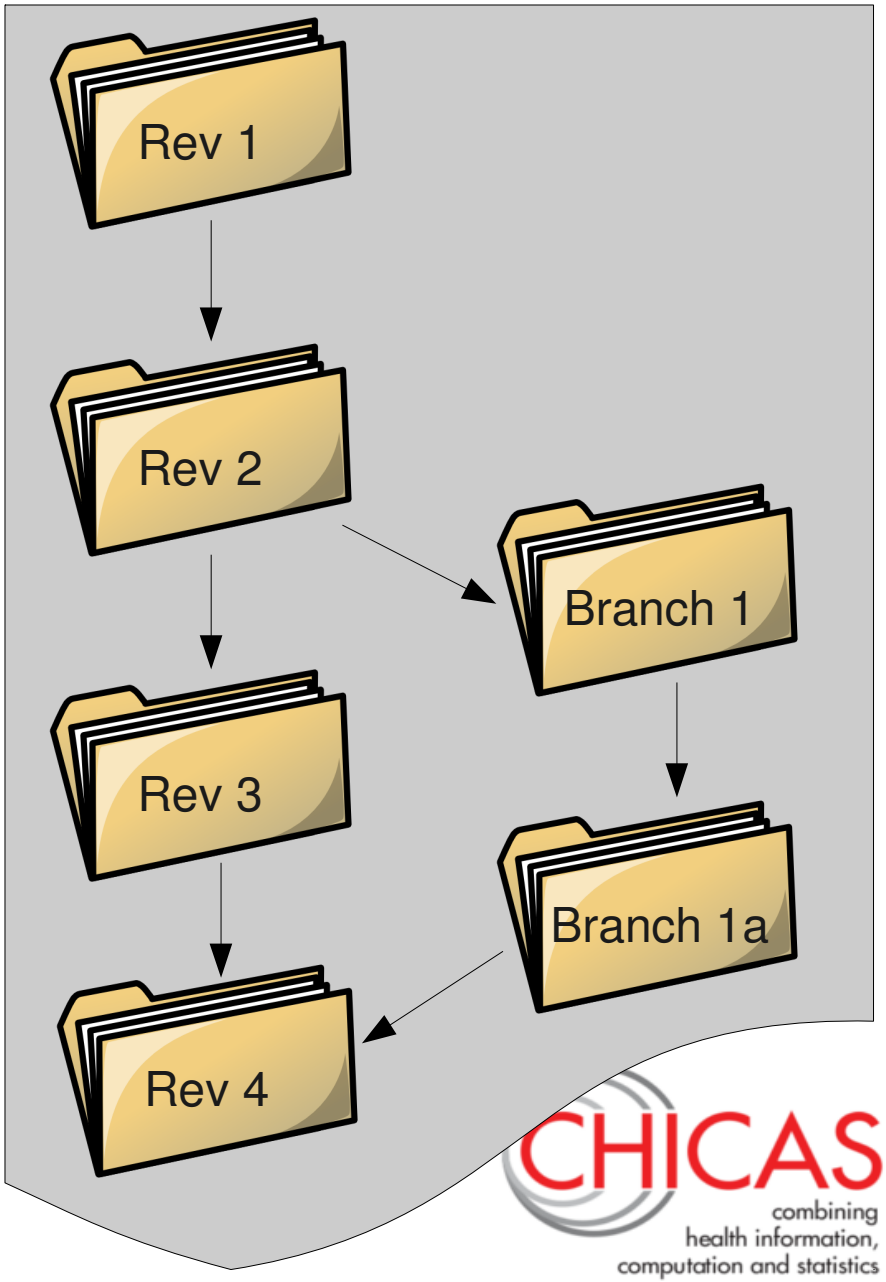
And so on...



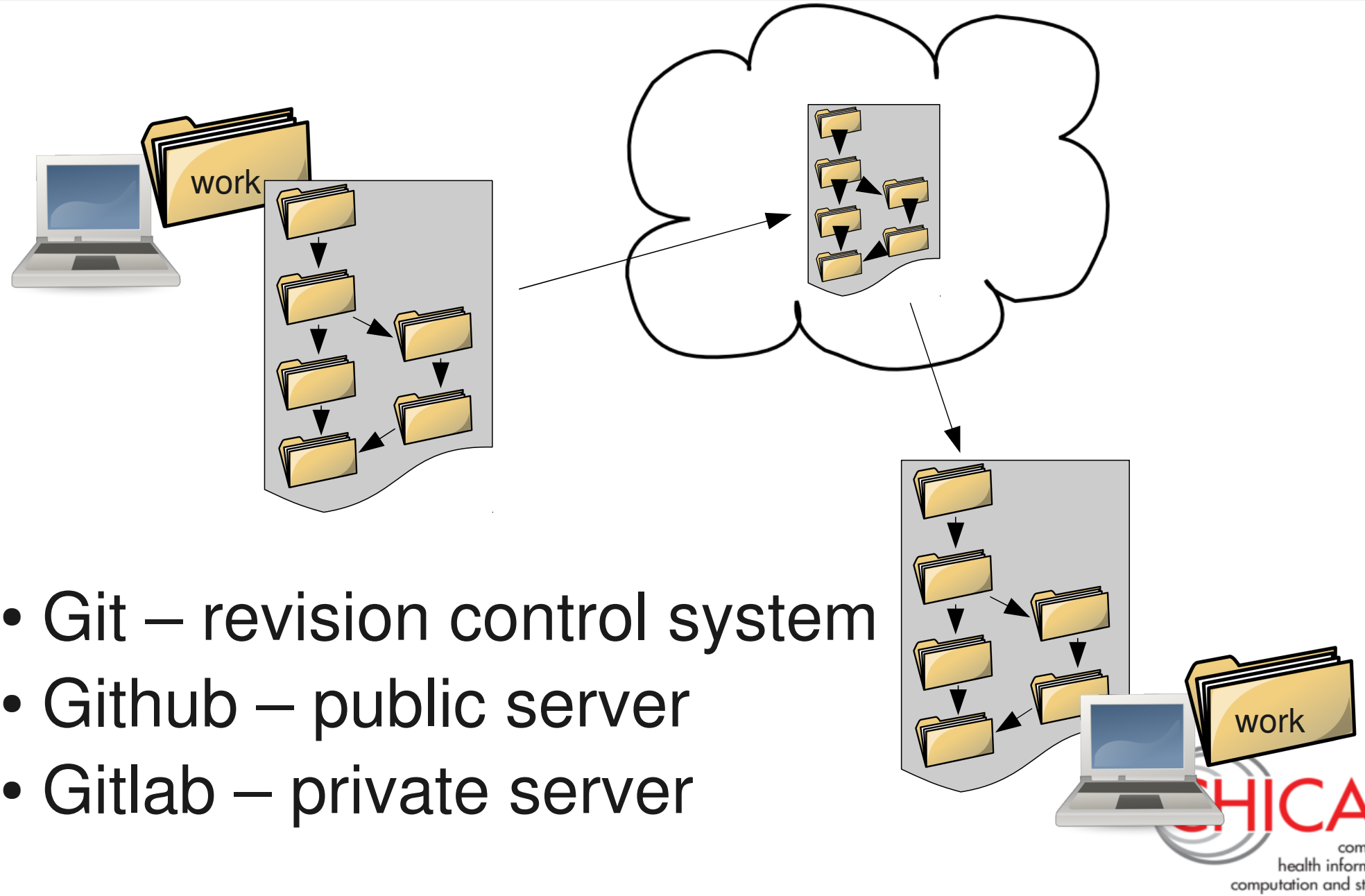
LANCASTER UNIVERSITY

CHICAS
combining
health information,
computation and statistics

branching/merging



collaborating



- Git – revision control system
- Github – public server
- Gitlab – private server

CHICAS
combining
health information,
computation and statistics

Time to move on

CHICAS
combining
health information,
computation and statistics