

Introduction to the tidyverse

Lars Hinrichs

English Department, The University of Texas at
Austin

2020/09/07 (updated: 2020-09-08)

tidyverse

Physically, what is the tidyverse?

- A meta-package that is meant to be loaded at the beginning of a work session.
- It actually loads a long list of dependency packages under the hood.
- Together, they change the state and some basic aspects of working with R.

Let's load the tidyverse

```
library(tidyverse)
```

```
## — Attaching packages —
```

```
## ✓ ggplot2 3.3.2      ✓ purrr    0.3.4
## ✓ tibble   3.0.3      ✓ dplyr    1.0.0
## ✓ tidyr    1.1.0      ✓ stringr  1.4.0
## ✓ readr    1.3.1      ✓forcats  0.5.0
```

```
## — Conflicts —
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

What changes if you load
tidyverse?

(1) Pipes

package: magrittr



The package name is a pun on *pipe*. (Actual title: La trahison des images; painted 1929)

Pipes are a way of programming.

Compare

```
order(fruitData$Juice, decreasing=TRUE)
```

to

```
fruitData$Juice %>% order(decreasing=TRUE)
```

Both commands do exactly the same thing.

In the abstract:

base-R:

```
functionname(arg1, arg2, ...)
```

tidyverse-R:

```
arg1 %>%  
  functionname(arg2, ...)
```

Advantage of coding with pipes

Code is much more legible than in base-R, e.g.

```
df %>%
  select(1:6) %>%
  na.omit() %>%
  unique()
```

and not

```
unique(na.omit(df[1:6,]))
```

Can be more easily ported, applied to different objects.

(2) Verbs, adverbs, nouns...

main functions: 'verbs'

wrapper functions that help run other functions (verbs) in a certain way: 'adverbs'

Verbs at the heart of tidyverse

The following verbs are all from the arch-package dplyr, which forms the center of the tidyverse.

`select()`

`filter()`

`summarise()`

`count()`

`arrange()`

(3) The 'tidy' aspect

Tidy data

Tidy datasets are easy to manipulate, model and visualise, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table.

Hadley Wickham (2014)

An untidy dataset

cars	buyers_gender	gender_percent	avg_hp
audi a4	male	50	120
golf gti	male	80	140
new beetle	male	40	90
audi a8	male	60	190

A tidy dataset

cars	make	model_suffix	buyers_gender	hp
a4	audi	avant g-tron	female	120
a8	audi	allroad quattro	NA	140
a8	audi	NA	female	90
a4	audi	limousine	male	190
a4	audi	NA	NA	140
golf	vw	gti	male	140
golf	vw	gte	female	90