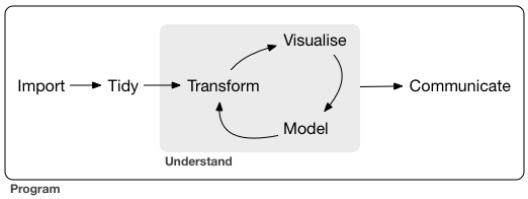
Welcome to the Tidyverse

What is the Tidyverse?



Because **R** is open-source, rather than a product released by a company such as SAS or Matlab, it relies heavily on user-created packages for its evolution.

The biggest revolution and modernization of **R** has come in the form of a collection of packages known as the **Tidyverse**. The father of the Tidyverse, and author of its most famous packages, is Hadley Wickham. However, many prominent statisticians and data scientists are now active contributors, including: Charlotte Wickham, Jenny Bryan, Yihui Xie, David Robinson, Julia Silge, and many many more.

This week, you learned "Base R": everything we did in today's activity did not require any extra packages. From now on, you will be learning everything through the lens of the Tidyverse.

Example

Consider the following dataset:

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Suppose we wish to accomplish the following edits to this data:

1. Drop the variables 'qsec' and 'vs', because we don't need them.
2. Include only automatic cars ('am' = 1)
3. Round the miles per gallon and weight of the cars to the nearest whole number.
4. Define a "fast car" to be one with high gear, cylinders, or horsepower. Make a variable for this, and count up how many cars are fast.

Here is how we would do it in Base R:

```
data(mtcars)

mtcars <- mtcars[, -c(7,8)]

auto <- mtcars$am == 1
mtcars <- mtcars[auto, ]

mtcars$mpg <- round(mtcars$mpg)
mtcars$wt <- round(mtcars$wt)

mtcars$speed <- rep("slow", nrow(mtcars))

fast_cars <- (mtcars$gear == 5 | mtcars$cyl == 8 | mtcars$hp > 200)

mtcars$speed[fast_cars] <- "fast"

mtcars$speed <- factor(mtcars$speed)
summary(mtcars$speed)
```

```
## fast slow
##    5    8
```

Here is the Tidy version:

```
data(mtcars)

mtcars <- mtcars %>%
  select(-qsec, -vs) %>%
  filter(am == 1) %>%
  mutate_at(
    vars(mpg, wt),
    funs(round)
  ) %>%
  mutate(
    speed = case_when(
      gear == 5 | cyl == 8 | hp > 200 ~ "fast",
      TRUE ~ "slow"
    )
  )

mtcars %>% count(speed)
```

```
## # A tibble: 2 x 2
##   speed     n
##   <chr> <int>
## 1 fast      5
## 2 slow      8
```

Advantages

The primary reasons to learn R via the Tidyverse, in my opinion:

- **Readability**: Code in the tidy style can be read more easily like an English sentence.
- **Reproducibility**: When the flow is clearer, it is easier to build sequences that can be re-applied to new data.
- **Presentation**: Tidy tools emphasize nice clean output. (RMarkdown is considered part of the Tidyverse!)
- **Culture:** The guiding principles behind Tidyverse packages emphasize sharing of code and data, responsible use of statistics (reproducibility!), and collaboration.
- **Community**: The Tidyverse R community is very active, diverse and welcoming. It is easy to get involved, and even the big names are happy to help beginners learn.