Data Manipulation using dplyr

Phil Canham
Corporate Insight Hub
October 2018

London Borough of Barking & Dagenham
lbbd.gov.uk

The DPLYR package

Dplyr is a language for data manipulation.

Most wrangling puzzles can be solved with knowledge of just 6 dplyr verbs (6 functions).

These will be the subject of this session.

Key dplyr verbs

select
filter
mutate
arrange
group_by
summarise

The piping operator

%>%

Shortcut: Ctrl/Shift M

The piping operator

%>%

This operator means THEN
Shortcut: Ctrl/Shift M

This operator lets us chain events together, in order and really helps to simplify quite complex sets of tasks

Think of a recipe: There are a series of steps

Boiling an egg

1. Bring the water to a **boil in a pan**.
2. Reduce the water to a rapid simmer.
3. Add the **egg**.
4. **Cook** the **egg** for 5 to 7 minutes.
5. Cool the **egg** slightly.
6. Remove the top off the **egg**.
7. Eat while the **egg** is warm!

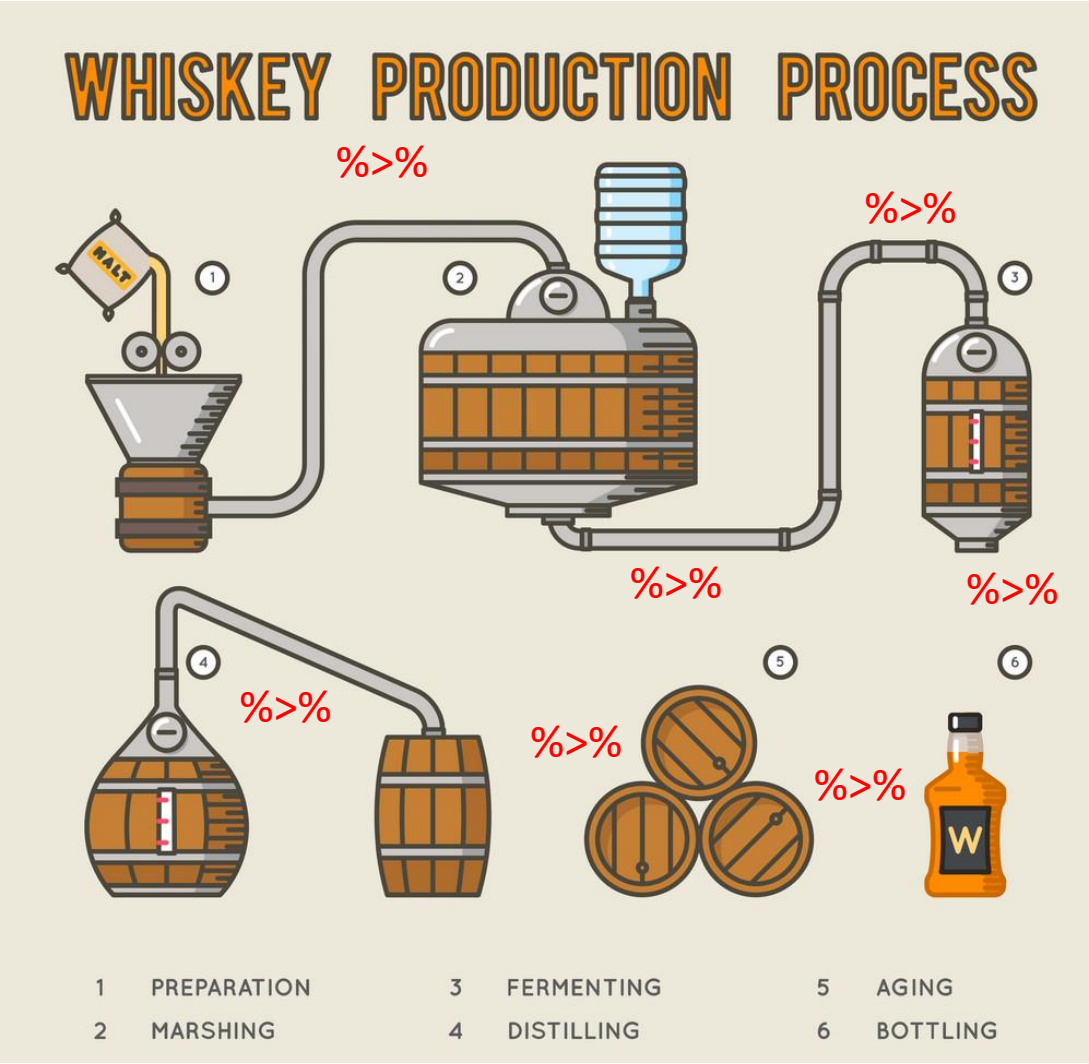
Boiling an egg

Bring the water to a **boil in a pan** **THEN**
Reduce the water to a rapid simmer. **THEN**
Add the **egg**. **THEN**
Cook the **egg** for 5 to 7 minutes. **THEN**
Cool the **egg** slightly. **THEN**
Remove the top off the **egg**. **THEN**
Eat while the **egg** is warm!

Boiling an egg

boil **water in a pan** %>%
Reduce the water to a rapid simmer. %>%
Add the **egg**. %>%
Cook the **egg** for 5 to 7 minutes. %>%
Cool the **egg** slightly. %>%
Remove the top off the **egg**. %>%
Eat while the **egg** is warm!

Whisky Production



And onto piping dplyr verbs

Grab a data_frame call “mammals”

```
mammals %>%
    mutate(mass_to_length = adult_body_mass_g / adult_head_body_len_mm) %>%
    arrange(desc(mass_to_length)) %>%
    select(species, mass_to_length)
```

```
## # A tibble: 5,416 x 2
##    species                mass_to_length
##    <chr>                           <dbl>
##  1 Balaena mysticetus              6539.
##  2 Balaenoptera musculus           5063.
##  3 Megaptera novaeangliae          2334.
##  4 Eschrichtius robustus           2309.
##  5 Balaenoptera physalus           2302.
##  6 Elephas maximus                 1704.
##  7 Eubalaena glacialis             1654.
##  8 Eubalaena australis             1625.
##  9 Balaenoptera edeni              1444.
## 10 Balaenoptera borealis           1203.
## # ... with 5,406 more rows
```