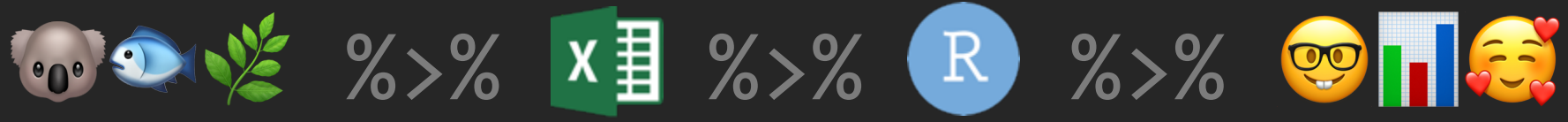
```
data_wrangling() &&
("manipulation" %in% R)
```



> day[3]

Ruan van Mazijk

Notes & slides will go up here:

tinyurl.com/r-with-ruan

(But I encourage you to make your own notes!)

`> workshop$outline[1:3]`

DAY 1

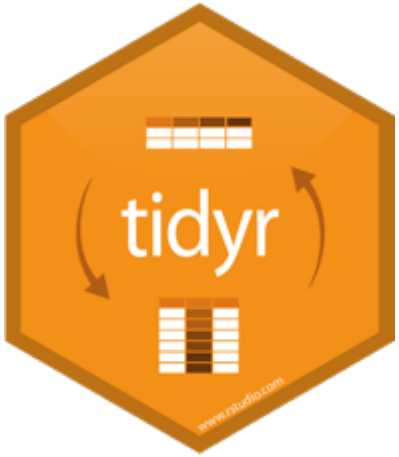
Tidy data principles
& `tidyr`

DAY 2

Manipulating data
& an intro to `dplyr`

DAY 3

Extending your data
with `mutate()`,
`summarise()`
& friends



`> workshop$outline[2:3]`

DAY 2

Manipulating data
& an intro to `dplyr`

DAY 3

Extending your data
with `mutate()`,
`summarise()`
& friends

dplyr::


```
# Verbs to manipulate your data

select()    # operates on columns
filter()    # operates on rows
```

```
data %>%
```

```
data %>%
  gather(key = veg_type, value = fix) %>%
```

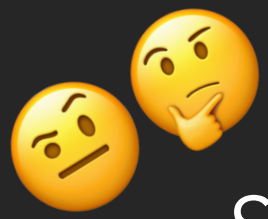
```
data %>%
  gather(key = veg_type, value = fix) %>%
  separate(fix, into = c("lon", "lat")) %>%
```

```
data %>%
  gather(key = veg_type, value = fix) %>%
  separate(fix, into = c("lon", "lat")) %>%
  select(veg_type, lon, lat, soil, plant_height) %>%
```

```r
data %>%
  gather(key = veg_type, value = fix) %>%
  separate(fix, into = c("lon", "lat")) %>%
  select(veg_type, lon, lat, soil, plant_height) %>%
  filter(plant_height %>% between(0.5, 10),
```

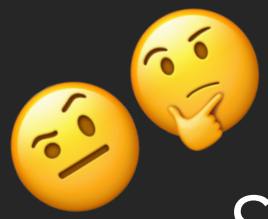
```
data %>%
  gather(key = veg_type, value = fix) %>%
  separate(fix, into = c("lon", "lat")) %>%
  select(veg_type, lon, lat, soil, plant_height) %>%
  filter(plant_height %>% between(0.5, 10),
         veg_type %in% c("fynbos",
                         "strandveld",
                         "renosterveld"))
```

```
data %>%
  gather(key = veg_type, value = fix) %>%
  separate(fix, into = c("lon", "lat")) %>%
  select(veg_type, lon, lat, soil, plant_height) %>%
  filter(plant_height %>% between(0.5, 10),
         veg_type %in% c("fynbos",
                         "strandveld",
                         "renosterveld"))
```



Summary statistics for each vegetation type?

```
data %>%
  gather(key = veg_type, value = fix) %>%
  separate(fix, into = c("lon", "lat")) %>%
  select(veg_type, lon, lat, soil, plant_height) %>%
  filter(plant_height %>% between(0.5, 10),
         veg_type %in% c("fynbos",
                         "strandveld",
                         "renosterveld")) %>%
  ???()
```



Summary statistics for each vegetation type?

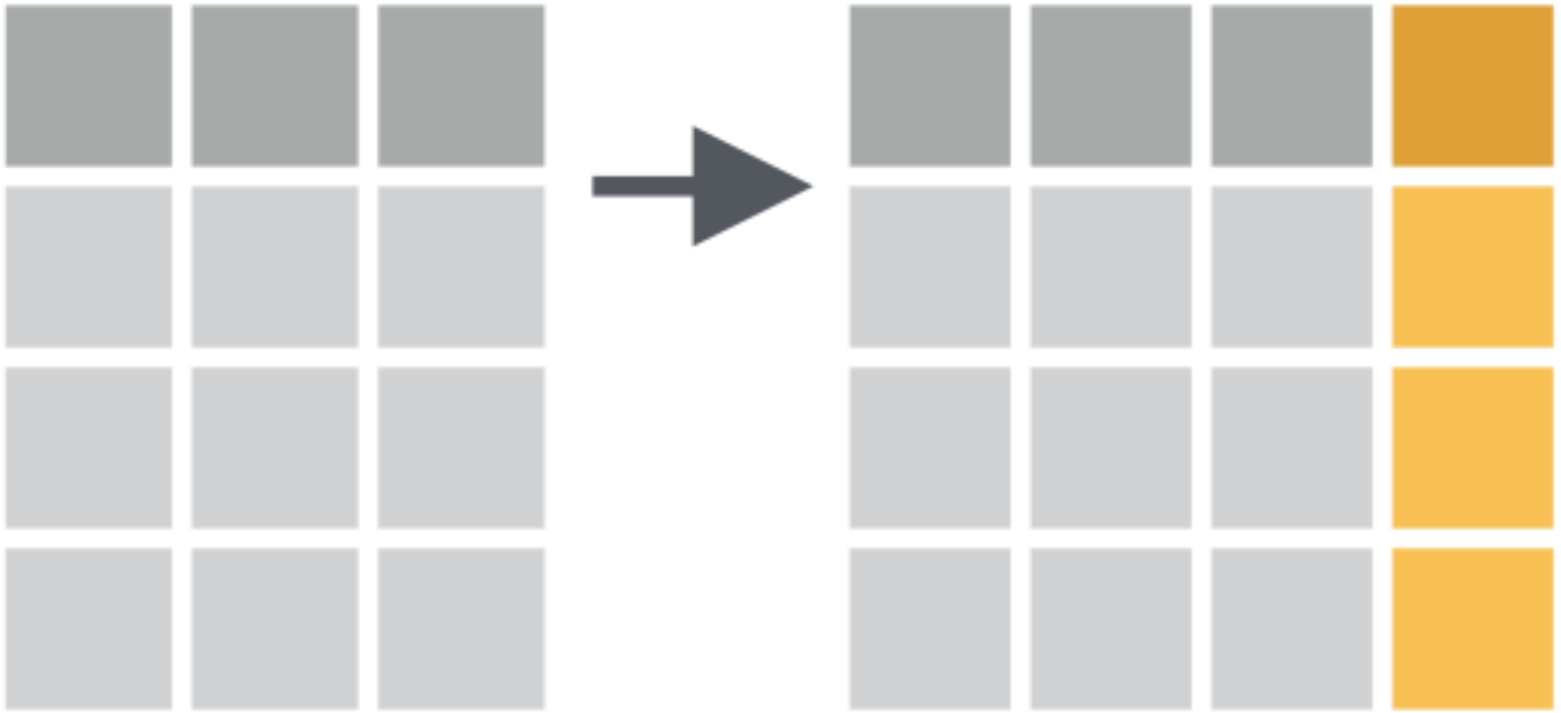
dplyr::


```
# Verbs to manipulate your data

select()    # operates on columns
filter()    # operates on rows
```

dplyr::


```
# Verbs to extend your data

mutate()    # operates on columns
group_by()  # operates on rows
summarise() # rows & columns
```

```
data %>%
  mutate(...)
```



CC BY SA RStudio https://www.rstudio.com/resources/cheatsheets/

```
data %>%
  mutate(...)
```

```
data %>%
  mutate(...)

data %>%
  mutate(BMI = height / weight)
```

```
data %>%
  mutate(...)

data %>%
  mutate(BMI = height / weight)

data %>%
  mutate(BMI = height / weight,
         BMI_std = scale(BMI))
```

```
data %>%
  mutate_all(...)
```

CC BY SA RStudio https://www.rstudio.com/resources/cheatsheets/

```
data %>%
  mutate_all(.funs, ...)

data %>%
  mutate_all(scale)

data %>%
  mutate_all(list(log, log1p))
```

```
data %>%
  mutate_if(.predicate, .funs)
```

CC BY SA RStudio https://www.rstudio.com/resources/cheatsheets/

```
data %>%
  mutate_if(.predicate, .funs, ...)

data %>%
  mutate_if(is.numeric, scale)

data %>%
  mutate_if(is.numeric, list(log, log1p))
```

dplyr::


```
# Verbs to extent your data

mutate()     # operates on columns
group_by()   # operates on rows
summarise()  # rows & columns
```

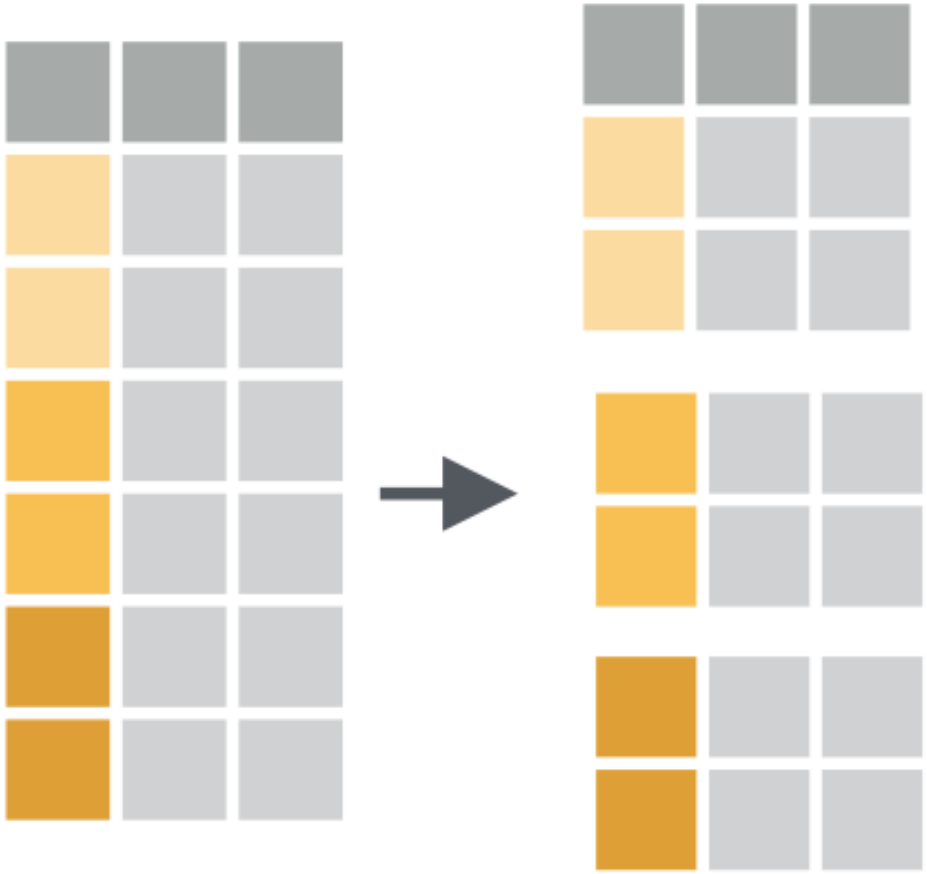
dplyr::


```
# Verbs to extent your data

mutate()     # operates on columns
group_by()   # operates on rows
summarise() # rows & columns
```

data

CC BY SA RStudio https://www.rstudio.com/resources/cheatsheets/

```
data %>% group_by(veg_type)
```



CC BY SA RStudio https://www.rstudio.com/resources/cheatsheets/

```
data %>% group_by(veg_type) %>% summarise(mean_plant_height =
                                          mean(plant_height))
```

CC BY SA RStudio https://www.rstudio.com/resources/cheatsheets/

```
data %>%
  group_by(veg_type) %>%
  summarise(mean_plant_height = mean(plant_height),
```

```
data %>%
  group_by(veg_type) %>%
  summarise(mean_plant_height = mean(plant_height),
            st_plant_height = sd(plant_height))
```

```r
data %>%
  group_by(veg_type) %>%
  summarise(mean_plant_height = mean(plant_height),
            st_plant_height = sd(plant_height))

data %>%
  group_by(veg_type) %>%
  summarise_if(is.numeric, mean)
```

```r
data %>%
  group_by(veg_type) %>%
  summarise(mean_plant_height = mean(plant_height),
            st_plant_height = sd(plant_height))

data %>%
  group_by(veg_type) %>%
  summarise_if(is.numeric, mean)

data %>%
  group_by(veg_type) %>%
  summarise_if(is.numeric, mean, na.rm = TRUE)
```

```r
data %>%
  group_by(veg_type) %>%
  summarise(mean_plant_height = mean(plant_height),
            st_plant_height = sd(plant_height))

data %>%
  group_by(veg_type) %>%
  summarise_if(is.numeric, mean)

data %>%
  group_by(veg_type) %>%
  summarise_if(is.numeric, mean, na.rm = TRUE)

data %>%
  group_by(veg_type) %>%
  summarise_if(is.numeric, list(mean, sd))
```

```
> demo()
```