### Socio-Informatics 348

Data Transformation with the Tidyverse Visualising the Transformations

#### Dr Lisa Martin

Department of Information Science Stellenbosch University

#### Practical Venues

- This morning, we tried out single venue based on attendance numbers
- Please rather go to the assigned venue (1.1, 1.2, etc.)

# RStudio Housekeeping

- Do not save workspace when closing RStudio
- Session ▷ Clear Workspace

#### Useful Visualisations

- Visualisations created by or adapted from Garrick Aden-Buie
- Useful for understanding how some dplyr transformations work

#### Mutate

	df	>
cat1	cat2	x
a	j	1
a	j	2
a	k	3
b	j	4
b	k	5
b	k	6
С	j	7
С	j	8
С	k	9

```
mutate(y = x - min(x))
   cat1 cat2
    a
    a
    a
    b
              6
```

```
df |>
  mutate(
    y = x - min(x)
)
```

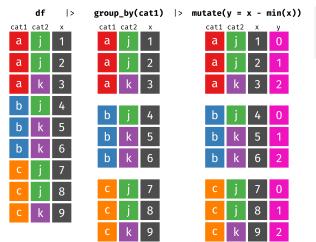
```
df|>
    mutate(
    y = x - min(x),
    after = cat2)

cat1 cat2    y    x

a    j    0    1

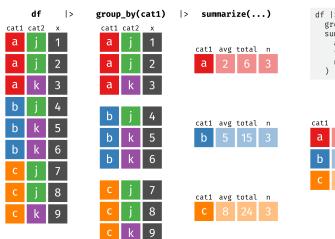
a    j    1   2
```

## Group and mutate



```
df |>
  group_by(cat1) |>
  mutate(
    y = x - min(x)
)
```

## Group and summarise I



```
df |>
  group_by(cat1) |>
  summarize(
  avg = mean(x),
  total = sum(x),
  n = n()
)
```

```
a 2 6 3
b 5 15 3
c 8 24 3
```

# Group and summarise II

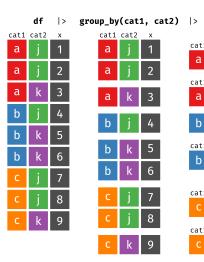


```
group_by(cat2) |> summarize(...)
                    cat2 avg total n
                    cat2 avg total n
```

```
df |>
  group_by(cat2) |>
  summarize(
   avg = mean(x),
   total = sum(x),
   n = n()
)
```

```
cat2 avg total n
j 4.4 22 5
k 5.75 23 4
```

# Group and summarise III

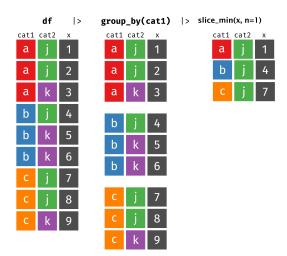


```
summarize(...) |>
      ungroup()
cat1 cat2 avg total n
cat1 cat2 avg total n
cat1 cat2 avg total
cat1 cat2 avg total
cat1 cat2 avg total
```

```
group_by(cat1, cat2) |>
summarize(
  avg = mean(x).
  total = sum(x),
  n = n()
) |>
ungroup()
cat1 cat2 avg total n
 b
```

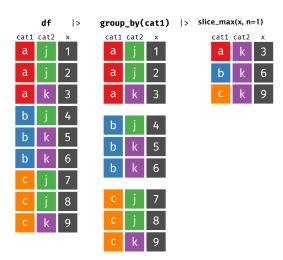
df |>

## Group and slice - min



df |> group\_by(cat1) |> slice\_min(x, n=1)

# Group and slice - max



df |> group\_by(cat1) |> slice\_max(x, n=1)