

R Data Analysis Course – Week 3

Grouping and Summarizing

Producing Useful Summaries

Ronnie Steinitz, Ph.D.
Dian Fossey Gorilla Fund

Grouping and Summarizing

- Collapse data into group-level summaries
 - means, counts, totals, etc.
- `group_by()` tells R how to split the data
 - e.g., by species, continent, or year
- `summarize()` defines what to compute for each group
 - e.g., `mean()`, `sd()`, `n()`
- Used together with the pipe (`%>%`) for clear, step-by-step logic

```
animals %>%  
  group_by(species) %>%  
  summarise(mean_speed = mean(speed_mps))
```

```
animals %>%  
  group_by(species) %>%  
  summarise(mean_speed = mean(speed_mps))
```

```
data = animals
```

species	speed_mps
Zebra	17
Lion	20
Hyena	14
Hyena	13
Gazelle	26
Lion	19
Zebra	18
Gazelle	23
Gazelle	25
Zebra	16
Lion	21
Gazelle	24
Hyena	15
Hyena	16

```
animals %>%
```

```
  group_by(species) %>%
```

```
  summarise(mean_speed = mean(speed_mps))
```

data = animals → group_by(species)

species	speed_mps
Zebra	17
Lion	20
Hyena	14
Hyena	13
Gazelle	26
Lion	19
Zebra	18
Gazelle	23
Gazelle	25
Zebra	16
Lion	21
Gazelle	24
Hyena	15
Hyena	16

species	speed_mps
Zebra	17
Zebra	18
Zebra	16

species	speed_mps
Gazelle	23
Gazelle	25
Gazelle	24
Gazelle	26

species	speed_mps
Lion	19
Lion	21
Lion	20

species	speed_mps
Hyena	15
Hyena	14
Hyena	13
Hyena	16

```
animals %>%
```

```
  group_by(species) %>%
```

```
  summarise(mean_speed = mean(speed_mps))
```

data = animals



group_by(species)



summarize(mean())

species	speed_mps
Zebra	17
Lion	20
Hyena	14
Hyena	13
Gazelle	26
Lion	19
Zebra	18
Gazelle	23
Gazelle	25
Zebra	16
Lion	21
Gazelle	24
Hyena	15
Hyena	16



species	speed_mps
Zebra	17
Zebra	18
Zebra	16

mean = 17

species	speed_mps
Gazelle	23
Gazelle	25
Gazelle	24
Gazelle	26

mean = 24.5

species	speed_mps
Lion	19
Lion	21
Lion	20

mean = 20

species	speed_mps
Hyena	15
Hyena	14
Hyena	13
Hyena	16

mean = 14.5

species	mean_speed
Gazelle	24.5
Hyena	14.5
Lion	20
Zebra	17

Auto-arranged by species!

```
animals %>%
```

```
  group_by(species) %>%
```

```
  summarise(mean_speed = mean(speed_mps),
```

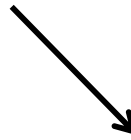
```
            max_speed = max(speed_mps),
```

```
            N = n())
```

data = animals



group_by(species)



summarize(mean(), max(), n())

species	speed_mps
Zebra	17
Lion	20
Hyena	14
Hyena	13
Gazelle	26
Lion	19
Zebra	18
Gazelle	23
Gazelle	25
Zebra	16
Lion	21
Gazelle	24
Hyena	15
Hyena	16



species	mean_speed	max_speed	N
Gazelle	24.5	26	4
Hyena	14.5	16	4
Lion	20	21	3
Zebra	17	18	3

Skills Learning

code-along lecture



Yellow-billed kite, Musanze 2025

Skills Application

laboratory exercise



Yellow-billed kite, Musanze 2025

Week 3 - Skills Application

Download **Skills Application Instructions** from [Google Drive folder > Week 3](#)

If you are [working on your own dataset](#):

- create a Markdown file which you will be using continuously and adding to each week.
- Name it: ***Lastname_Firstname_Data.Rmd*** in your **wd**.
Save **your own dataset** in your **wd**

If you are [working on palmer penguins raw dataset](#):

- create a Markdown file just for this week.
- Name it: ***Lastname_Firstname_Week3.Rmd*** in ***`/Week 3`***

In your new Markdown file: Use code blocks (Ctrl+Alt+i) to load packages, import dataset into the environment, and save data as an object (e.g., ***`data <-`***). Then follow *instructions*.