Data Wrangling

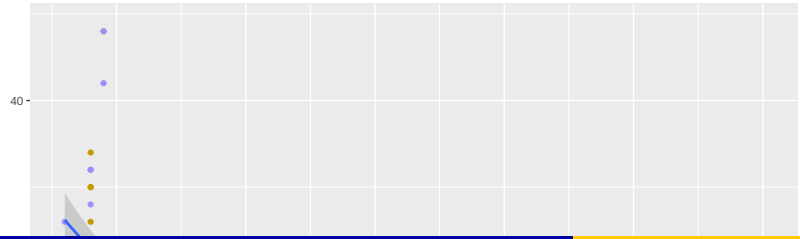
Data management and analysis in environmental sciences

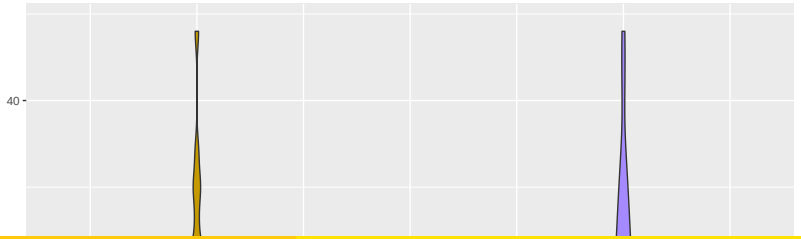
Jesse Radolinski, Johannes Ingrisch

Department of Ecology, University of Innsbruck

Last Week: Visualisation

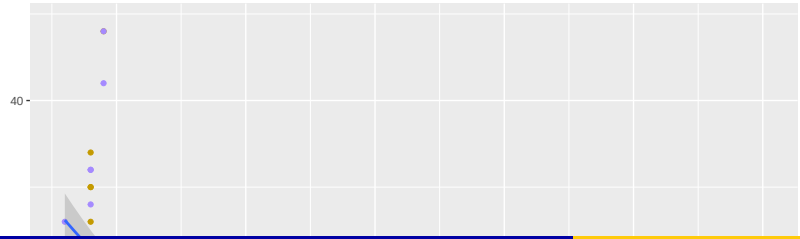

```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class)) +
  geom_smooth()
ggplot(data = mpg) +
  geom_violin(aes(x = class, y = hwy,
                  fill = class))
```

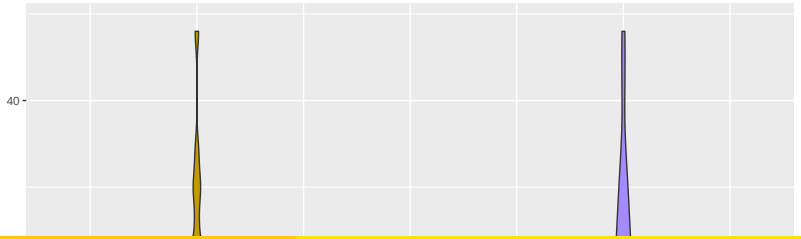




Last Week: Visualisation

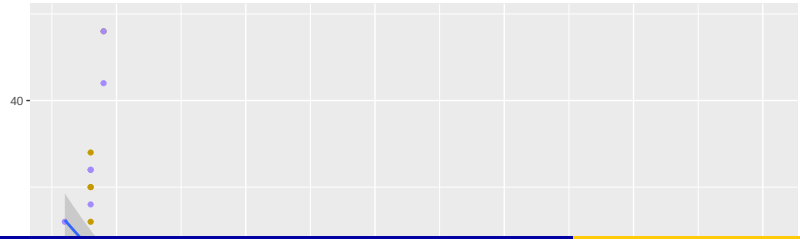

```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class)) +
  geom_smooth()
ggplot(data = mpg) +
  geom_violin(aes(x = class, y = hwy,
                  fill = class))
```

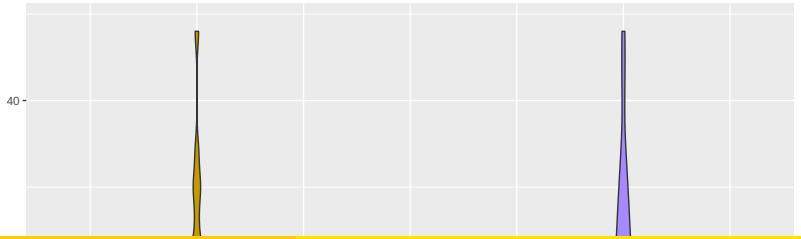




Last Week: Visualisation

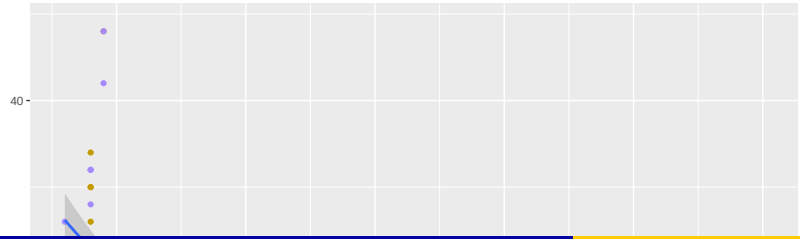

```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class)) +
  geom_smooth()
ggplot(data = mpg) +
  geom_violin(aes(x = class, y = hwy,
                  fill = class))
```

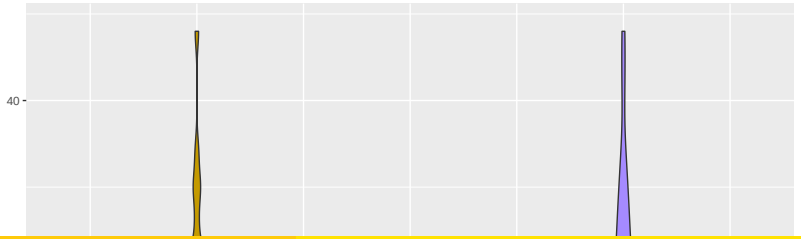




Last Week: Visualisation

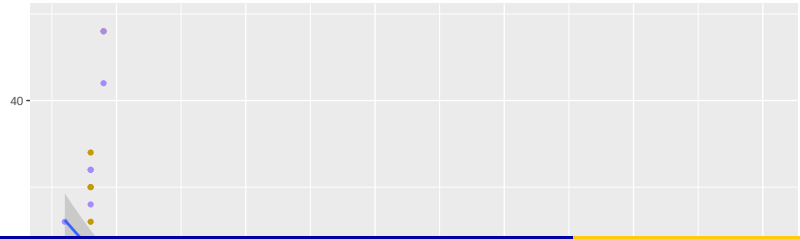

```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class)) +
  geom_smooth()
ggplot(data = mpg) +
  geom_violin(aes(x = class, y = hwy,
                  fill = class))
```

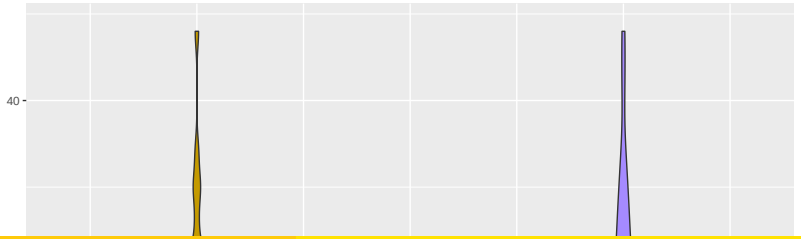




Last Week: Visualisation


```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class)) +
  geom_smooth()
ggplot(data = mpg) +
  geom_violin(aes(x = class, y = hwy,
                  fill = class))
```





The Pipe: %>%

R-Studio Shortcut: **Ctrl + Shift + M**

The Pipe: %>%

Passes object on the lefthand side as first argument of function on righthand side.

- `x %>% f(y)` is similar to `f(x,y)`

```
filter(mpg, class == "seater" | class == "compact")
```

Part 2

Summarize data: summarise()

Collapses data frame to a single row ->

```
mpg %>%
  summarise(hwy = mean(hwy))
```

```
# A tibble: 1 x 1
    hwy
  <dbl>
1  23.4
```

summarise()

Also possible for multiple functions and variables

```
mpg %>%
  summarise(hwy_avg = mean(hwy),
```

Part 2

Summarize data: summarise()

Collapses data frame to a single row ->

```
mpg %>%
  summarise(hwy = mean(hwy))
```

```
# A tibble: 1 x 1
    hwy
  <dbl>
1  23.4
```

summarise()

Also possible for multiple functions and variables

```
mpg %>%
  summarise(hwy_avg = mean(hwy),
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```

Reading Data

Reading text files

- most general function is read.table() from base-R

```
"dataset","x","y"
"dino",55.3846,97.1795
"dino",51.5385,96.0256
"dino",46.1538,94.4872
"dino",42.8205,91.4103
"dino",40.7692,88.3333
"dino",38.7179,84.8718
"dino",35.641,79.8718
"dino",33.0769,77.5641
"dino",28.9744,74.4872
```