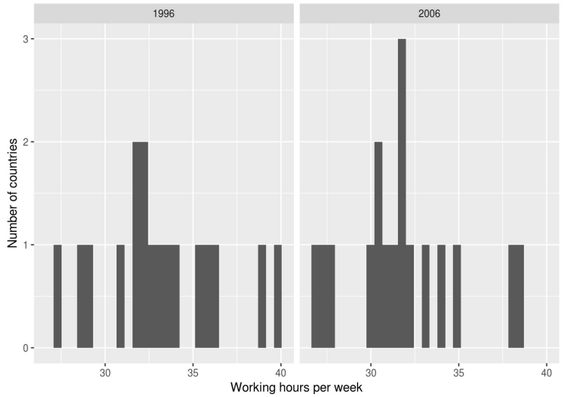
COMMUNICATING WITH DATA IN THE TIDYVERSE

Visualizing aspects of data with facets

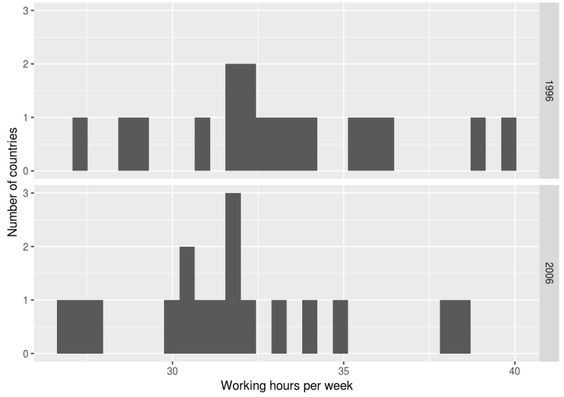
Timo Grossenbacher
Data Journalist

The facet_grid() function

```
ilo_data <- ilo_data %>%
  filter(year == "1996" |
  year == "2006")

ggplot(ilo_data) +
  geom_histogram(aes(
    x = working_hours)) +
  labs(x = "Working hours per week",
       y = "Number of countries") +
  facet_grid(. ~ year)
```

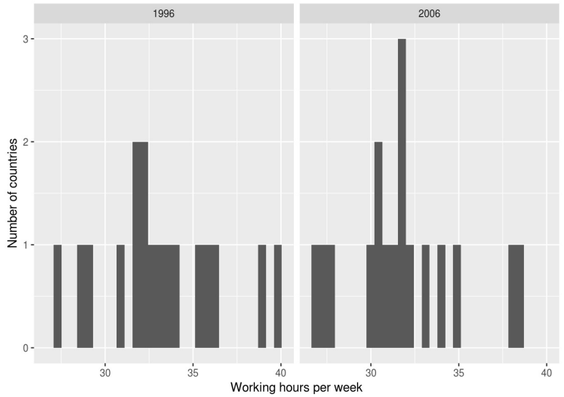



```
ilo_data <- ilo_data %>%
  filter(year == "1996" |
  year == "2006")

ggplot(ilo_data) +
  geom_histogram(aes(
    x = working_hours)) +
  labs(x = "Working hours per week",
       y = "Number of countries") +
  facet_grid(year ~ .)
```

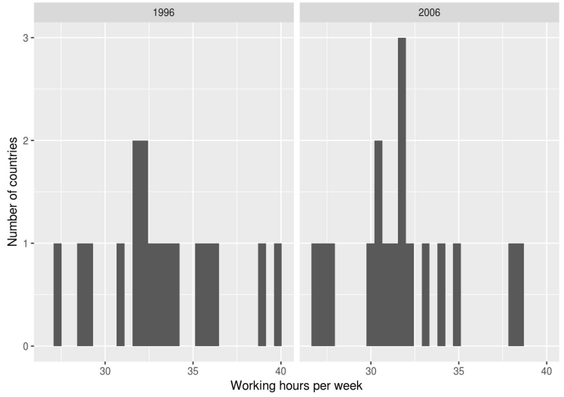


The facet_grid() function

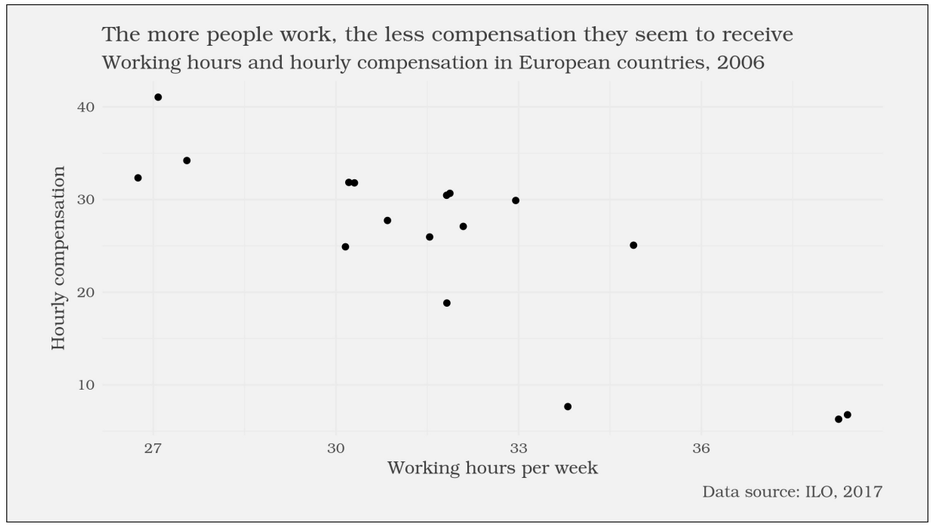
```
ilo_data <- ilo_data %>%
  filter(year == "1996" |
  year == "2006")

ggplot(ilo_data) +
  geom_histogram(aes(
    x = working_hours)) +
  labs(x = "Working hours per week",
       y = "Number of countries") +
  facet_grid(. ~ year)
```

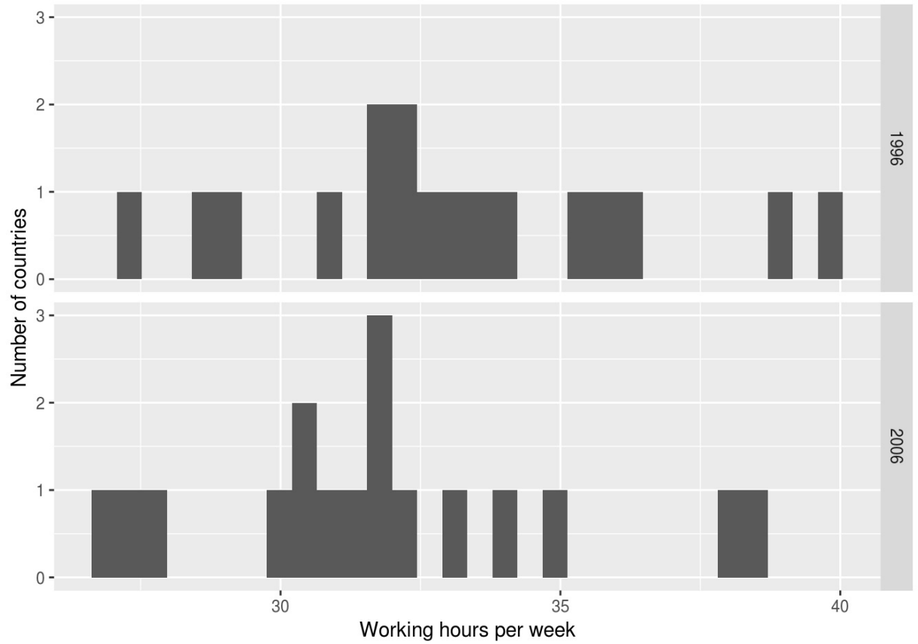



```
ilo_data <- ilo_data %>%
  filter(year == "1996" |
  year == "2006")

ggplot(ilo_data) +
  geom_histogram(aes(
    x = working_hours)) +
  labs(x = "Working hours per week",
       y = "Number of countries") +
  facet_wrap(facets = ~ year)
```



A faceted scatter plot





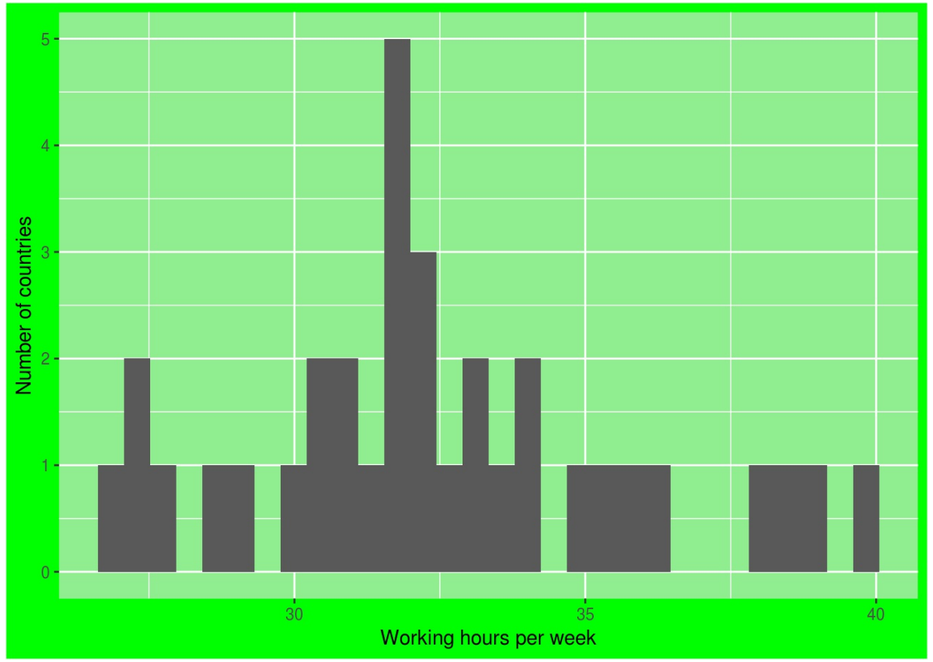
Styling faceted plots

```
strip.background
strip.text
...
```

Defining your own theme function

```
theme_green <- function(){
  theme(
    plot.background =
      element_rect(fill = "green"),
    panel.background =
      element_rect(fill =
        "lightgreen")
  )
}
```

```
ggplot(ilo_data) +
  geom_histogram(aes(
    x = working_hours)) +
  labs(x = "Working hours per week",
       y = "Number of countries") +

theme_green()
```



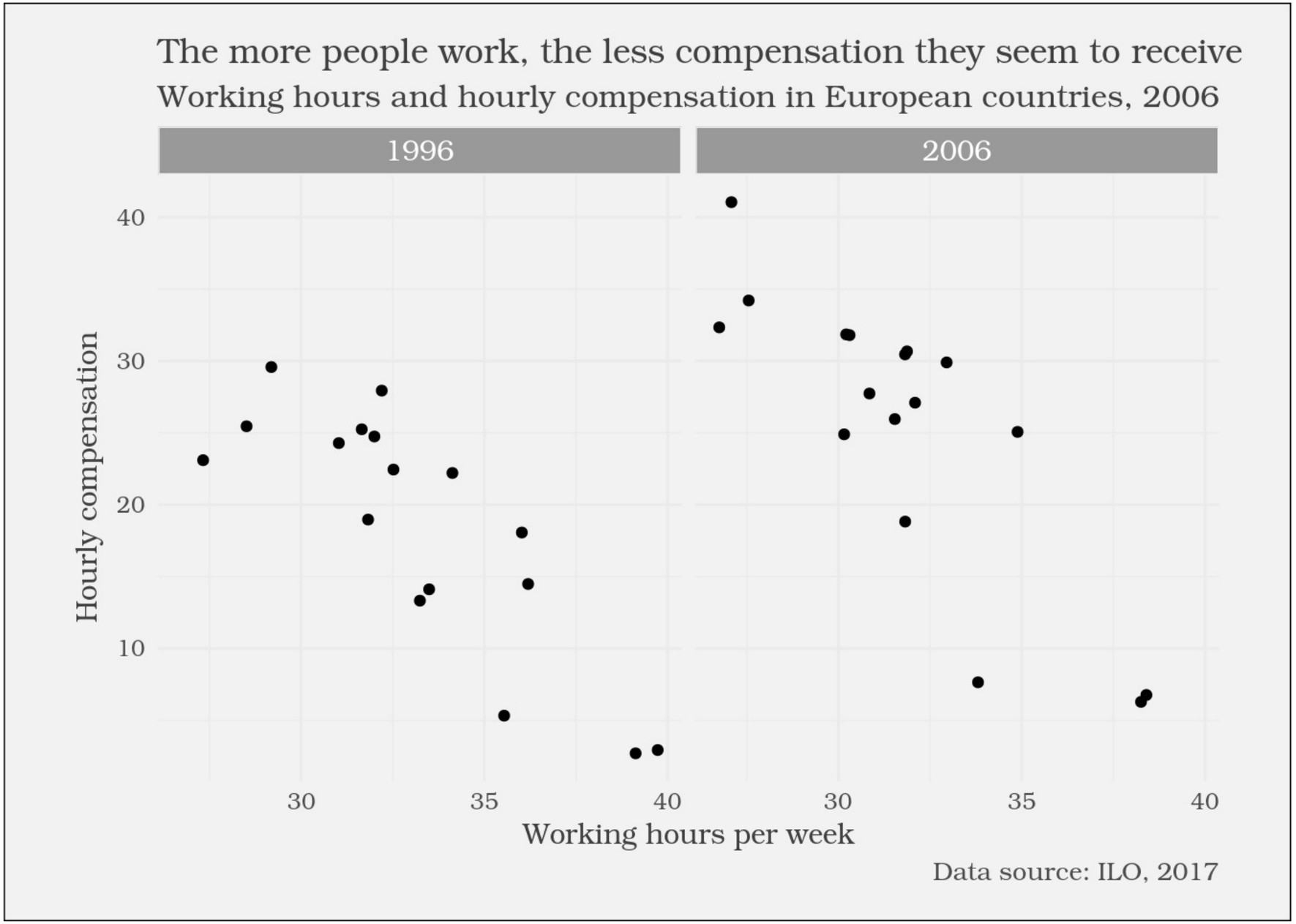
COMMUNICATING WITH DATA IN THE TIDYVERSE

Let's practice!

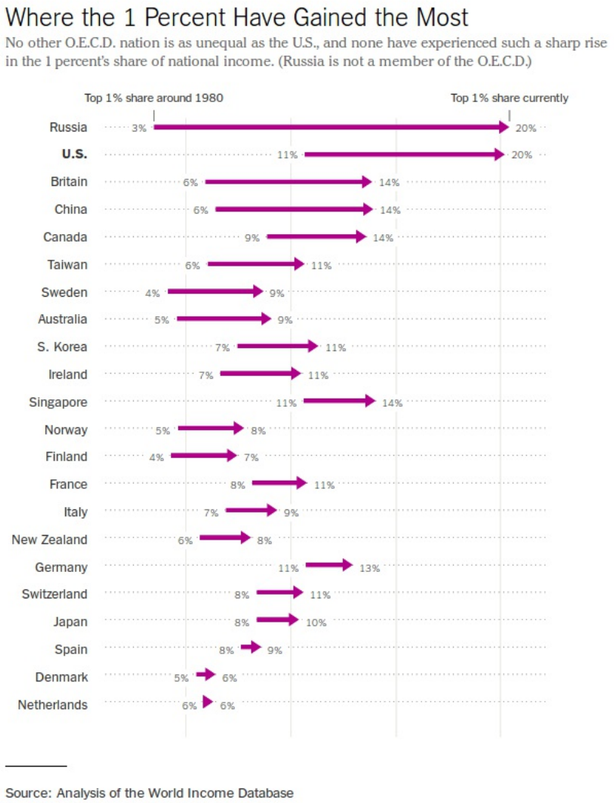
COMMUNICATING WITH DATA IN THE TIDYVERSE

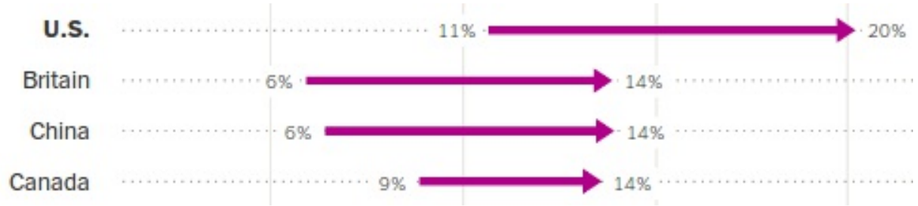
A custom plot to emphasize change

Timo Grossenbacher
Data Journalist


The more people work, the less compensation they seem to receive
Working hours and hourly compensation in European countries, 2006
1996
2006
Hourly compensation
40
30
20
10
30
35
40
30
35
40
Working hours per week
Data source: ILO, 2017

The dot plot

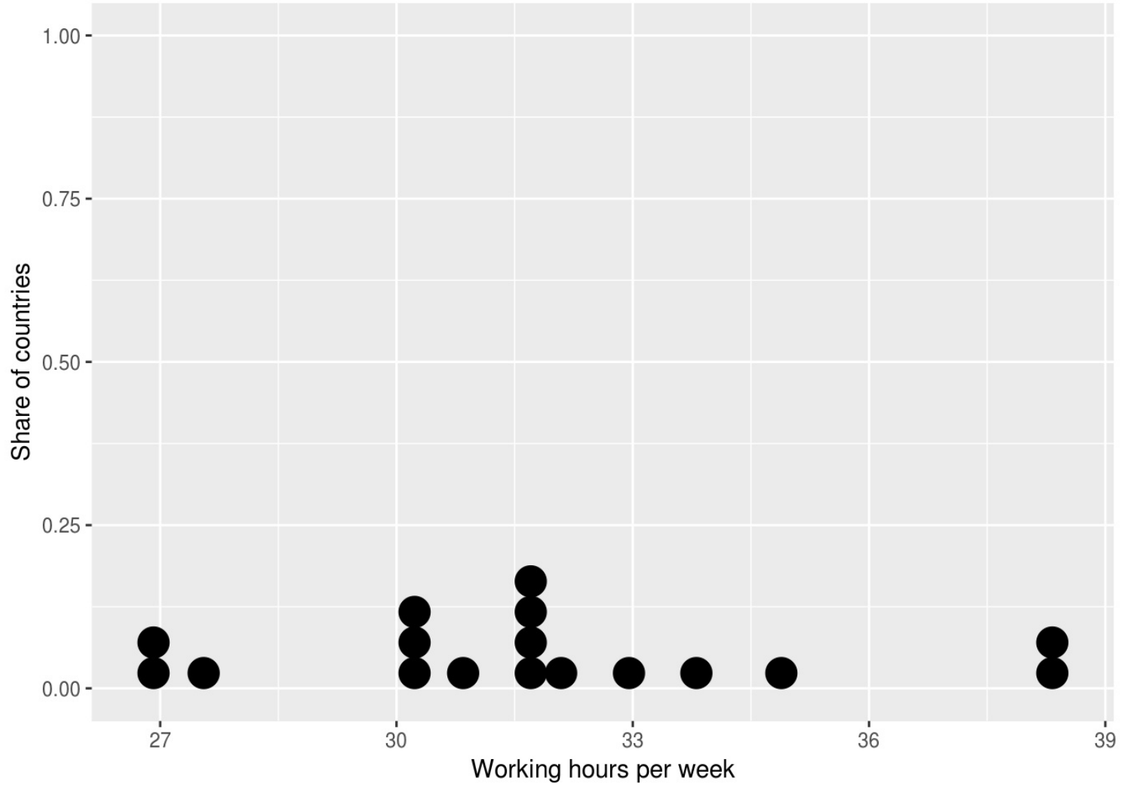




[1] New York Times (https://www.nytimes.com/2017/11/17/upshot/income-inequality-united-states.html){{0}}

Dot plots with ggplot2

```
ggplot((ilo_data %>% filter(year == 2006))) +
  geom_dotplot(aes(x = working_hours)) +
  labs(x = "Working hours per week",
       y = "Share of countries")
```



Dot plots with ggplot2: the geom_path() function

```
?geom_path
```

geom_path() *connects the observations in the order in which they appear in the data.*

```
ilo_data %>%
    arrange(country)
```

```
# A tibble: 34 x 4
     country   year hourly_compensation working_hours
      <fctr> <fctr>                <dbl>         <dbl>
1    Austria   1996                24.75      31.99808
2    Austria   2006                30.46      31.81731
3    Belgium   1996                25.25      31.65385
4    Belgium   2006                31.85      30.21154
5 Czech Rep.   1996                 2.94      39.72692
6 Czech Rep.   2006                 6.77      38.40000
# ... with 28 more rows
```

Dot plots with ggplot2: the geom_path() function

```
ggplot() +
    geom_path(aes(x = numeric_variable, y = numeric_variable))
```

```
ggplot() +
    geom_path(aes(x = numeric_variable, y = factor_variable))
```

```
ggplot() +
    geom_path(aes(x = numeric_variable, y = factor_variable),
                arrow = arrow(___))
```

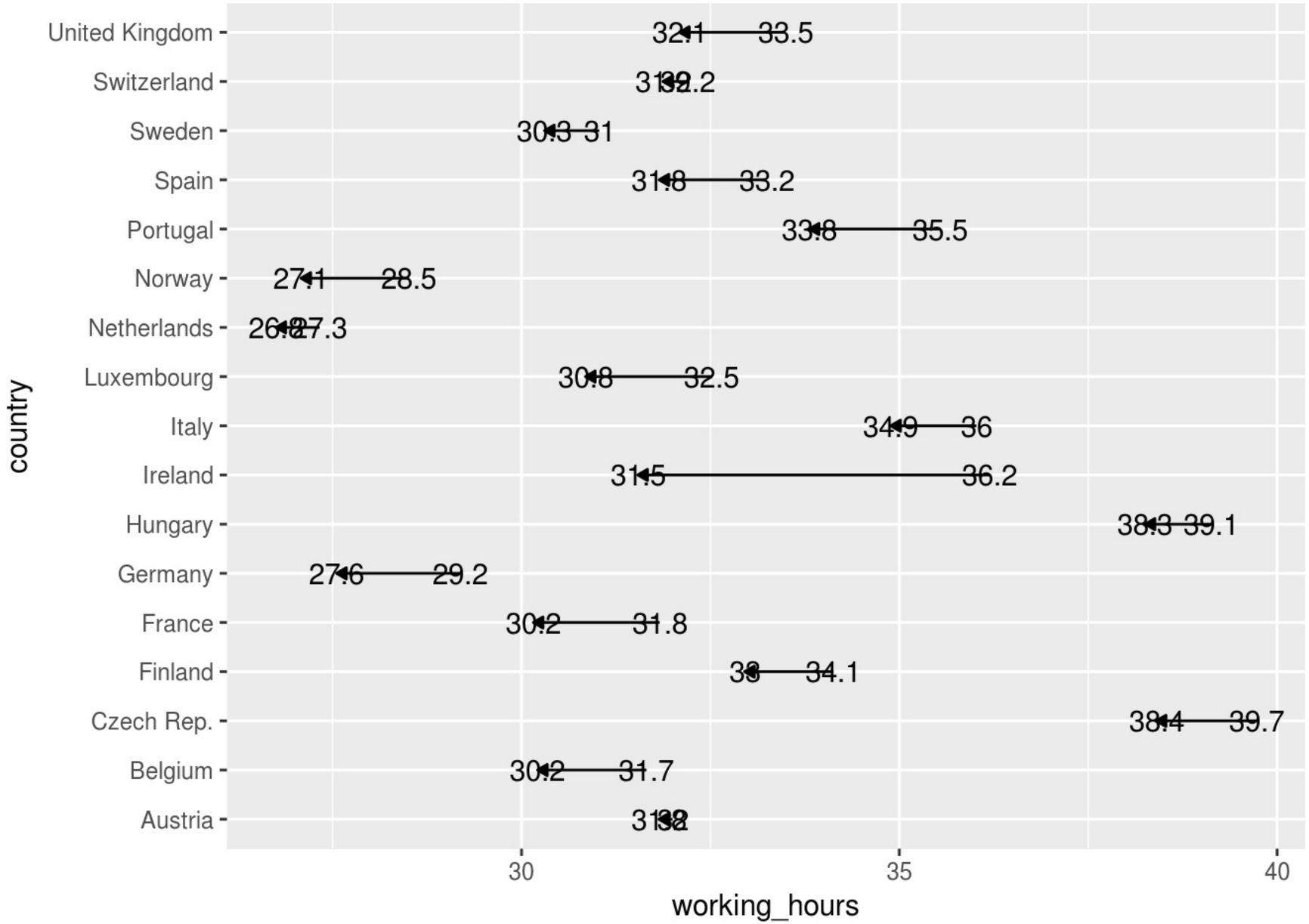
COMMUNICATING WITH DATA IN THE TIDYVERSE

Let's try out geom_path!

COMMUNICATING WITH DATA IN THE TIDYVERSE

Polishing the dot plot

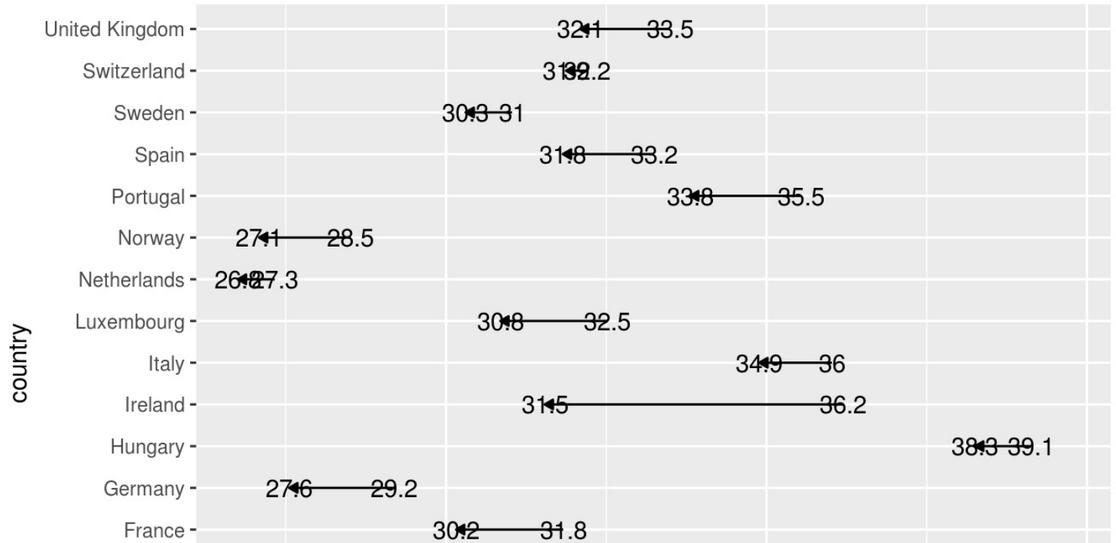
Timo Grossenbacher
Data Journalist


United Kingdom
Switzerland
Sweden
Spain
Portugal
Norway
Netherlands
Luxembourg
Italy
Ireland
Hungary
Germany
France
Finland
Czech Rep.
Belgium
Austria
country
30
35
40
working_hours

Factor levels

- The order of factor **levels** determine the order of appearance in ggplot2.

```
ilo_data$country

 [1] Austria        Belgium        Czech Rep.     Finland
 [5] France         Germany        Hungary    ...
 ...
 17 Levels: Austria Belgium Czech Rep. Finland France ... United Kingdom
```



Reordering factors with the forcats package

- Needs to be loaded with

  library(forcats)

- fct_drop for dropping levels
- fct_rev for reversing factor levels
- fct_reorder for reordering them.

[1] Learn more at tidyverse.org (http://forcats.tidyverse.org/)

The fct_reorder function

```
ilo_data

# A tibble: 34 x 4
      country   year hourly_compensation working_hours
       <fctr> <fctr>                <dbl>         <dbl>
 1    Austria   1996                24.75      31.99808
 2    Austria   2006                30.46      31.81731
 3    Belgium   1996                25.25      31.65385
 4    Belgium   2006                31.85      30.21154
 5 Czech Rep.   1996                 2.94      39.72692
 6 Czech Rep.   2006                 6.77      38.40000
```

```
ilo_data <- ilo_data  %>%
    mutate(country = fct_reorder(country, working_hours, mean))

ilo_data$country

 [1] Austria        Belgium        Czech Rep.     Finland
 [5] France         Germany        Hungary    ...
 ...
17 Levels: Netherlands Norway Germany Sweden ... Czech Rep.
```

The fct_reorder function

```
ilo_data <- ilo_data  %>%
    mutate(country = fct_reorder(country, working_hours, mean))
```

```
# A tibble: 34 x 4
       country   year hourly_compensation working_hours
        <fctr> <fctr>               <dbl>         <dbl>
 1    Austria   1996               24.75      31.99808
 2    Austria   2006               30.46      31.81731
 3    Belgium   1996               25.25      31.65385
 4    Belgium   2006               31.85      30.21154
 5 Czech Rep.   1996                2.94      39.72692
 6 Czech Rep.   2006                6.77      38.40000
```

mean(c(31.99808, 31.81731))

mean(c(31.65385, 30.21154))

mean(c(39.72692, 38.40000))

Nudging labels with hjust and vjust

```
ggplot(ilo_data) +
  geom_path(aes(...)) +
  geom_text(
        aes(...,
            hjust = ifelse(year == "2006",
              1.4,
              -0.4)
        )
    )
```

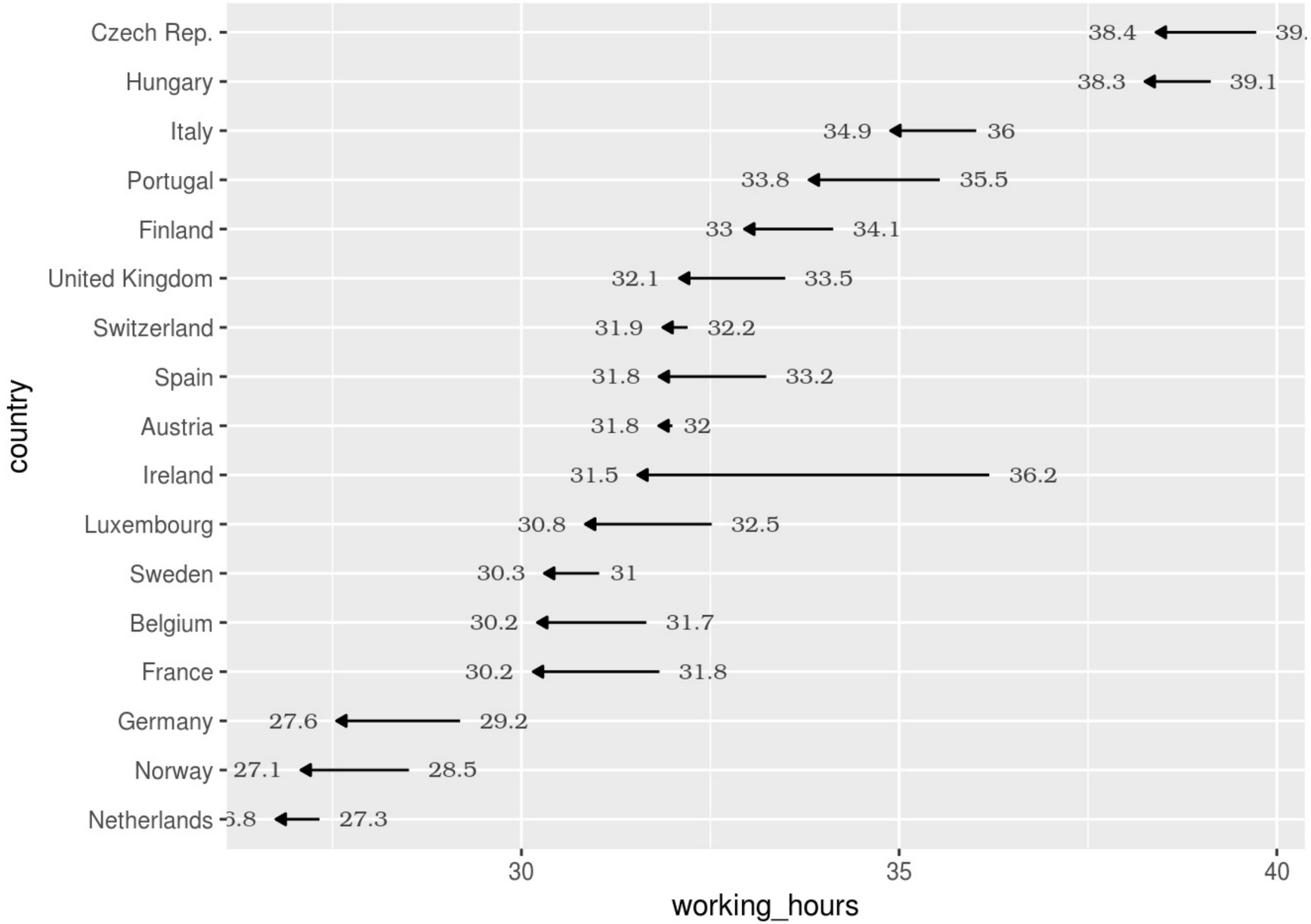
COMMUNICATING WITH DATA IN THE TIDYVERSE

Let's practice!

COMMUNICATING WITH DATA IN THE TIDYVERSE

Finalizing the plot for different audiences and devices

Timo Grossenbacher
Data Journalist

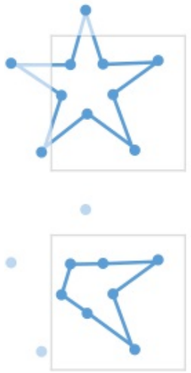

Czech Rep.
Hungary
Italy
Portugal
Finland
United Kingdom
Switzerland
Spain
Austria
Ireland
Luxembourg
Sweden
Belgium
France
Germany
Norway
Netherlands
38.4
39.
38.3
39.1
34.9
36
33.8
35.5
33
34.1
32.1
33.5
31.9
32.2
31.8
33.2
31.8
32
31.5
36.2
30.8
32.5
30.3
31
30.2
31.7
30.2
31.8
27.6
29.2
27.1
28.5
27.3
country
30
35
40
working_hours

coord_cartesian vs. xlim / ylim

```
ggplot_object +
    coord_cartesian(xlim = c(0, 100), ylim = c(10, 20))
```

```
ggplot_object +
    xlim(0, 100) +
    ylim(10, 20)
```

coord_cartesian vs. xlim / ylim



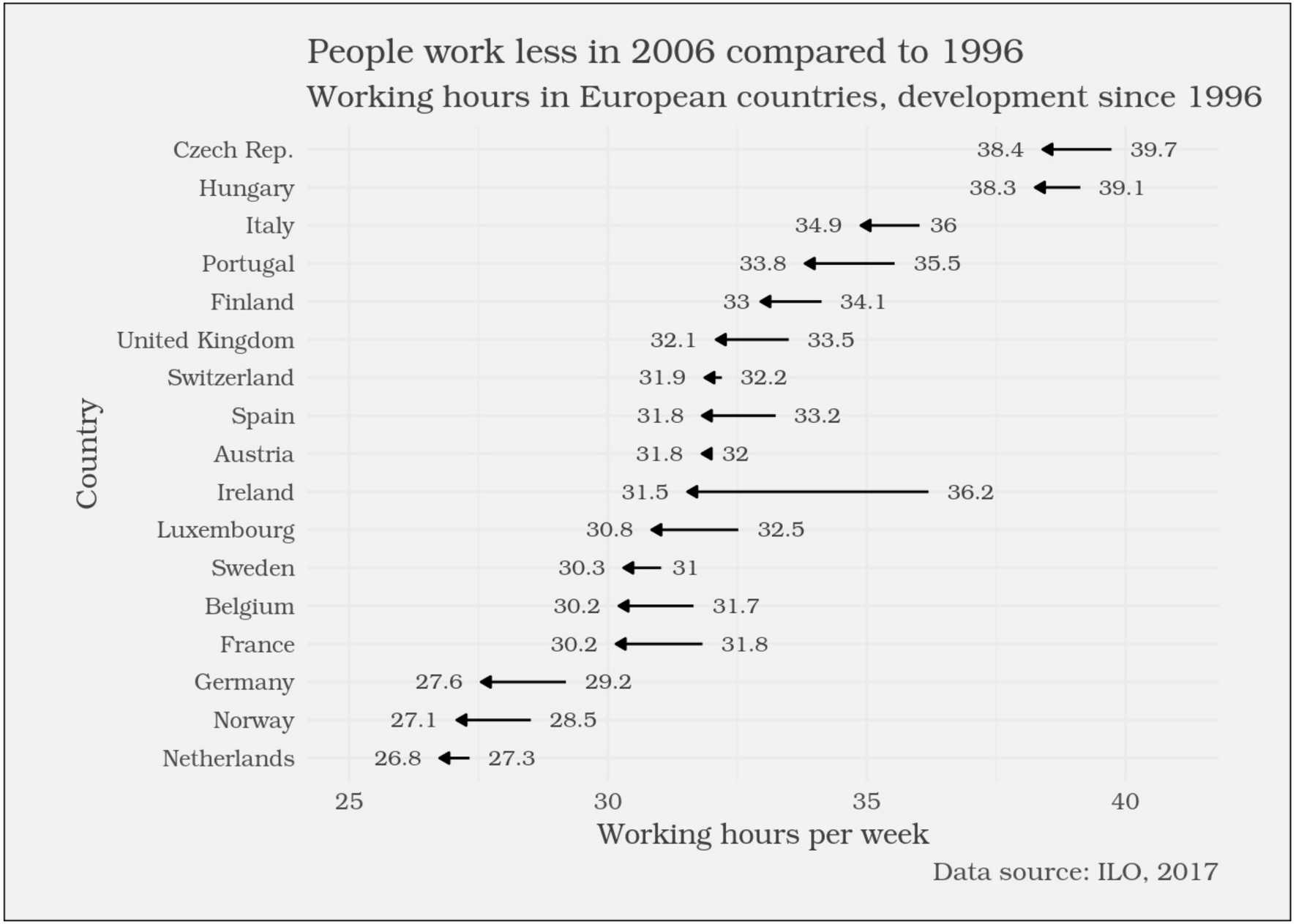
Without clipping (preferred)

```
t + coord_cartesian(
xlim = c(0, 100), ylim = c(10, 20))
```

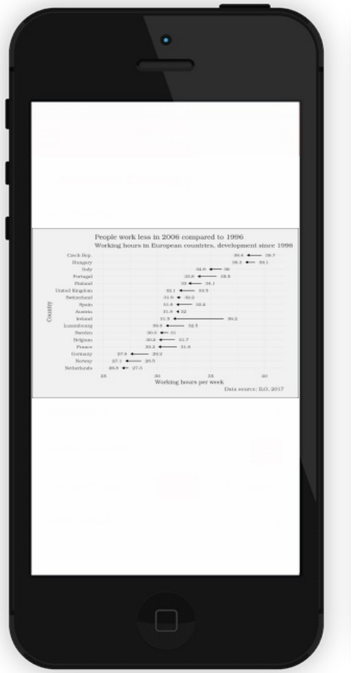
With clipping (removes unseen data points)

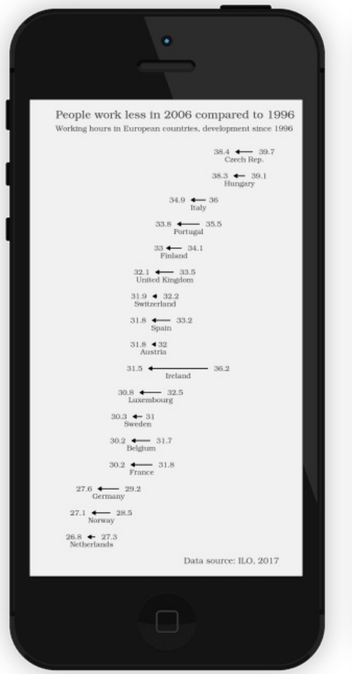
```
t + xlim(0, 100) + ylim(10, 20)
```

[1] Taken from RStudio Data Visualization Cheat Sheet (https://github.com/rstudio/cheatsheets/raw/master/data-visualization-2.1.pdf)


People work less in 2006 compared to 1996
Working hours in European countries, development since 1996
Country
Czech Rep.
Hungary
Italy
Portugal
Finland
United Kingdom
Switzerland
Spain
Austria
Ireland
Luxembourg
Sweden
Belgium
France
Germany
Norway
Netherlands
38.4
39.7
38.3
39.1
34.9
36
33.8
35.5
33
34.1
32.1
33.5
31.9
32.2
31.8
33.2
31.8
32
31.5
36.2
30.8
32.5
30.3
31
30.2
31.7
30.2
31.8
27.6
29.2
27.1
28.5
26.8
27.3
25
30
35
40
Working hours per week
Data source: ILO, 2017

Desktop vs. Mobile audiences





COMMUNICATING WITH DATA IN THE TIDYVERSE

Let's produce these plots!