

Customizing graphs with R and ggplot2

Data Visualization Meetup Munich

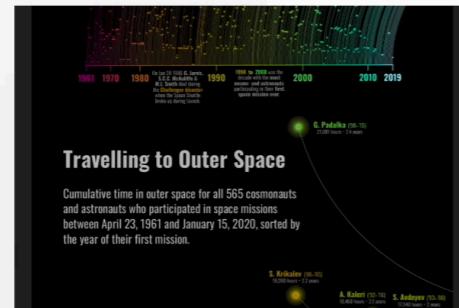


What happened so far

WHAT I APPRECIATE IN DATAVIZ CHALLENGES

Inspiration

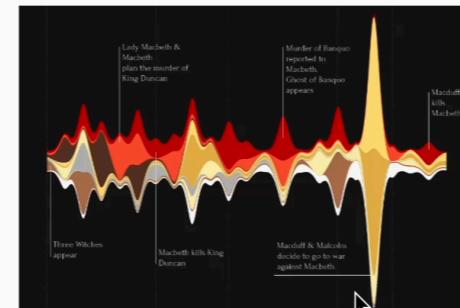
Be inspired by what others share
Learn tricks from others
Improve your skills



Source: Cédric Scherer ([Github](#))

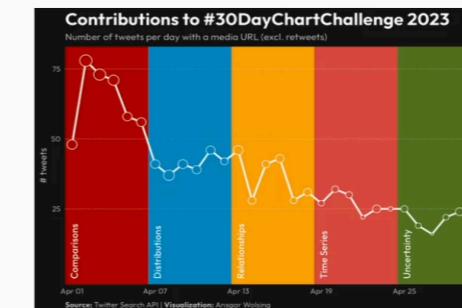
Experimentation

Get out of your comfort zone, but in a safe way
New chart types, tools, topics



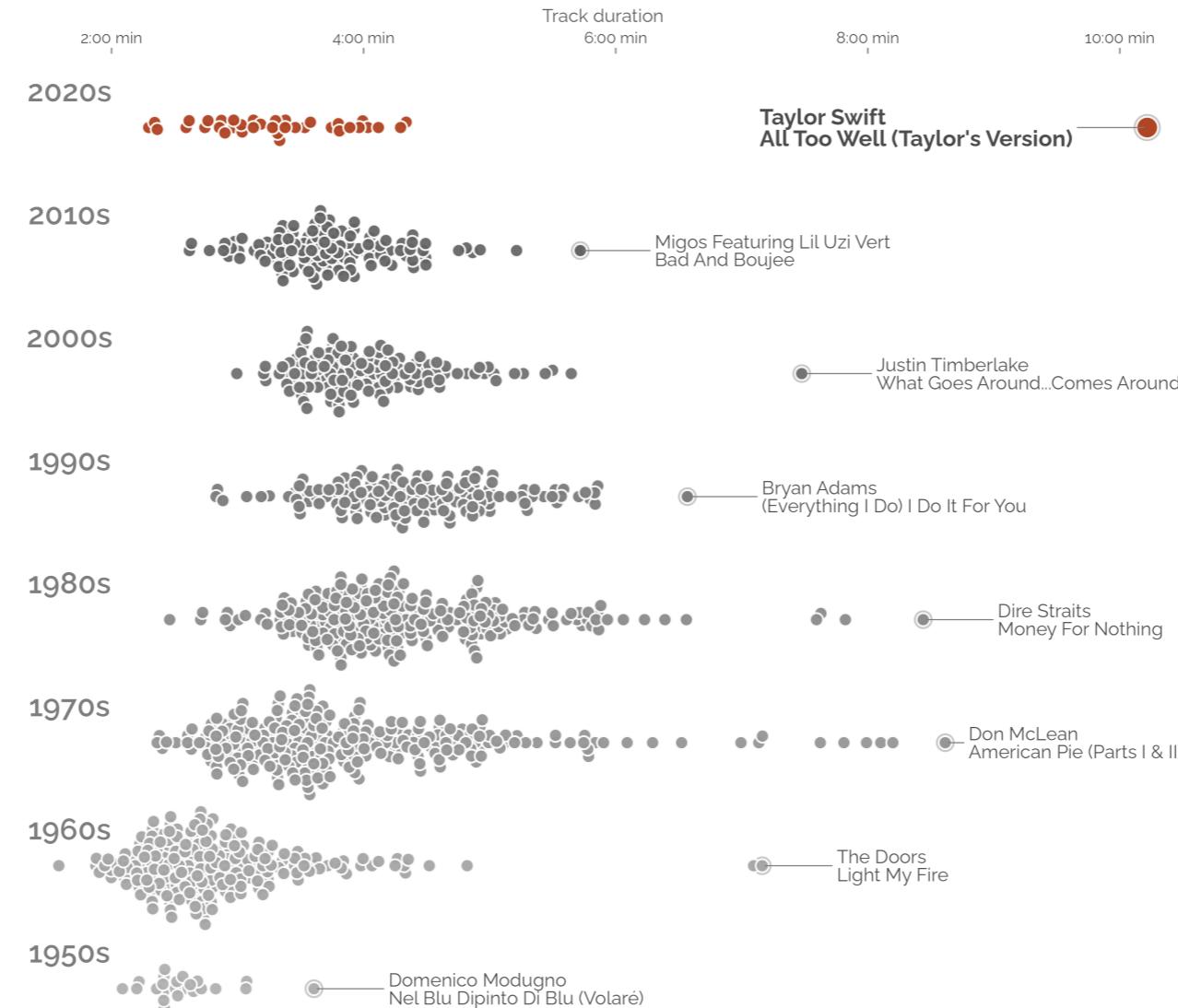
Connections

Connect with like-minded people from across the globe
Exposure & recognition
Feedback & support



"All Too Well" is the Longest Song in U.S. Chart History

Taylor Swift's new version of "All Too Well" is the longest track to ever top the Billboard Hot 100 with a duration of 10 minutes and 13 seconds. The song replaces Don McLean's song "American Pie" (8 minutes and 36 seconds), which became no. 1 on 15 January 1972, after nearly 50 years. Each no. 1 song is a dot. The longest no. 1 song in each decade is highlighted.



Source: Billboard Hot 100 (Kaggle), Spotify API | Visualization: Ansgar Wolsing

Data Visualization Meetup Munich | July 8th, 2025 | Ansgar Wolsing



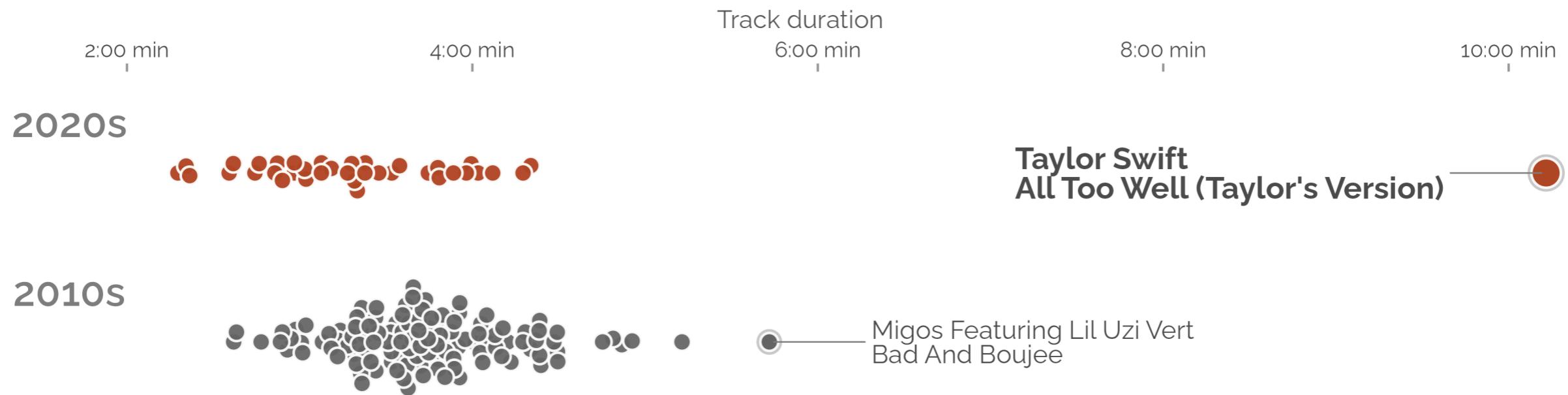
"All Too Well" is the Longest Song in U.S. Chart History

Taylor Swift's new version of "All Too Well" is the longest track to ever top the Billboard Hot 100 with a duration of 10 minutes and 13 seconds. The song replaces **Don McLean's song "American Pie"** (8 minutes and 36 seconds), which became no. 1 on 15 January 1972, after nearly 50 years. Each no. 1 song is a dot. The longest no. 1 song in each decade is highlighted.



"All Too Well" is the Longest Song in U.S. Chart History

Taylor Swift's new version of "All Too Well" is the longest track to ever top the Billboard Hot 100 with a duration of 10 minutes and 13 seconds. The song replaces **Don McLean's song "American Pie"** (8 minutes and 36 seconds), which became no. 1 on 15 January 1972, after nearly 50 years. Each no. 1 song is a dot. The longest no. 1 song in each decade is highlighted.



Inspiration



How Taylor Swift Outsmarted Her Record Label

 **Rick Beato** 
5,07 Mio. Abonnenten

 Abonniert 

 38.922



 Teilen

 Herunterladen

...

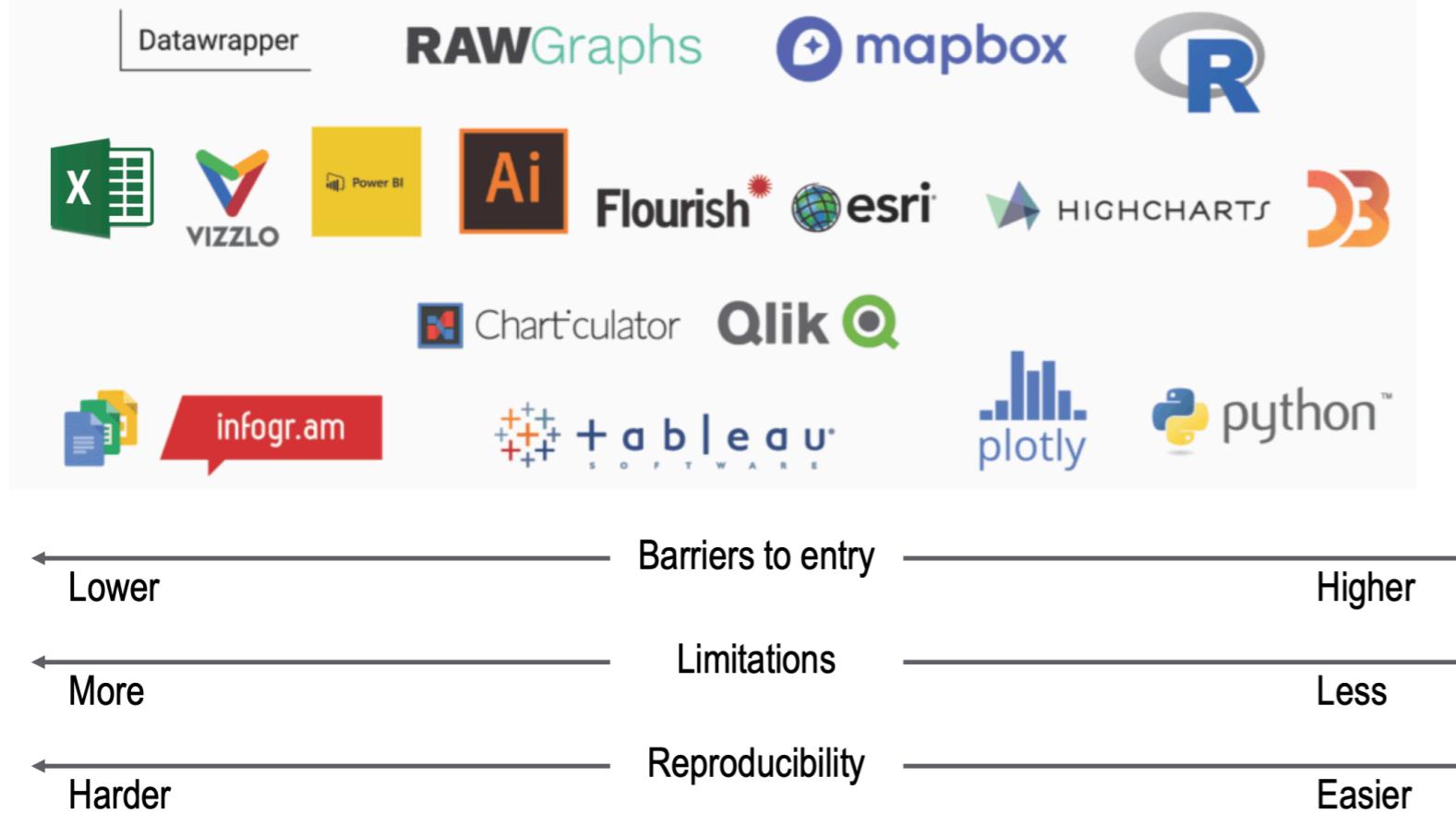
1 Mio. Aufrufe vor 3 Jahren

In this episode we discuss how Taylor Swift is owning the music industry and beating the system through her album re-records.

Holiday Sale: COUPON CODE: RB1200 ...mehr



There's a continuum of dataviz tools



Source: Schwabish (2021)



Why creating charts with code?

Reproducibility

Every step is explicit
and repeatable

Customization

fine-tune, create non-standard charts, share standards
programmatically

Automation

create functions for repeated tasks



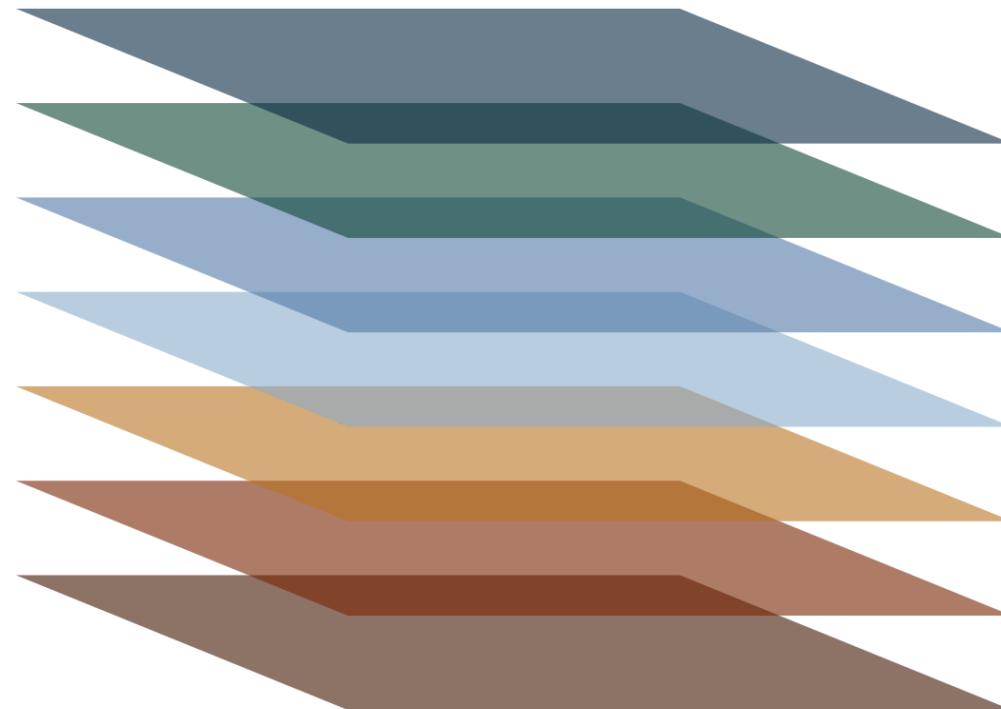
ggplot2 at a glance

Implementation of the Grammar of Graphics (Wilkinson 2005) in R ("gg")



ggplot2 at a glance

Implementation of the Grammar of Graphics (Wilkinson 2005) in R ("gg")



Theme
Coordinate system
Facets
Scales
Layers
Aesthetics
Data



Let's dive into the code



Load the required libraries

```
1 library(tidyverse)  
2 library(ggtext)
```

`library()` loads R packages into the R session.
`{ggplot2}` is part of the `{tidyverse}`

artist_song	artist	song	first_no1_date	decade	duration_s
Britney Spears - ...Baby One More Time	Britney Spears	...Baby One More Time	1999-01-30	1990	211.066
Nelson - (Can't Live Without Your) Love And Affection	Nelson	(Can't Live Without Your) Love And Affection	1990-09-29	1990	236.435
Bryan Adams - (Everything I Do) I Do It For You	Bryan Adams	(Everything I Do) I Do It For You	1991-07-27	1990	394.133
B.J. Thomas - (Hey Won't You Play) Another Somebody Done Somebody Wrong Song	B.J. Thomas	(Hey Won't You Play) Another Somebody Done Somebody Wrong Song	1975-04-26	1970	206.066
The Rolling Stones - (I Can't Get No) Satisfaction	The Rolling Stones	(I Can't Get No) Satisfaction	1965-07-10	1960	222.813
Cutting Crew - (I Just) Died In Your Arms	Cutting Crew	(I Just) Died In Your Arms	1987-05-02	1980	280.400



Data preparation

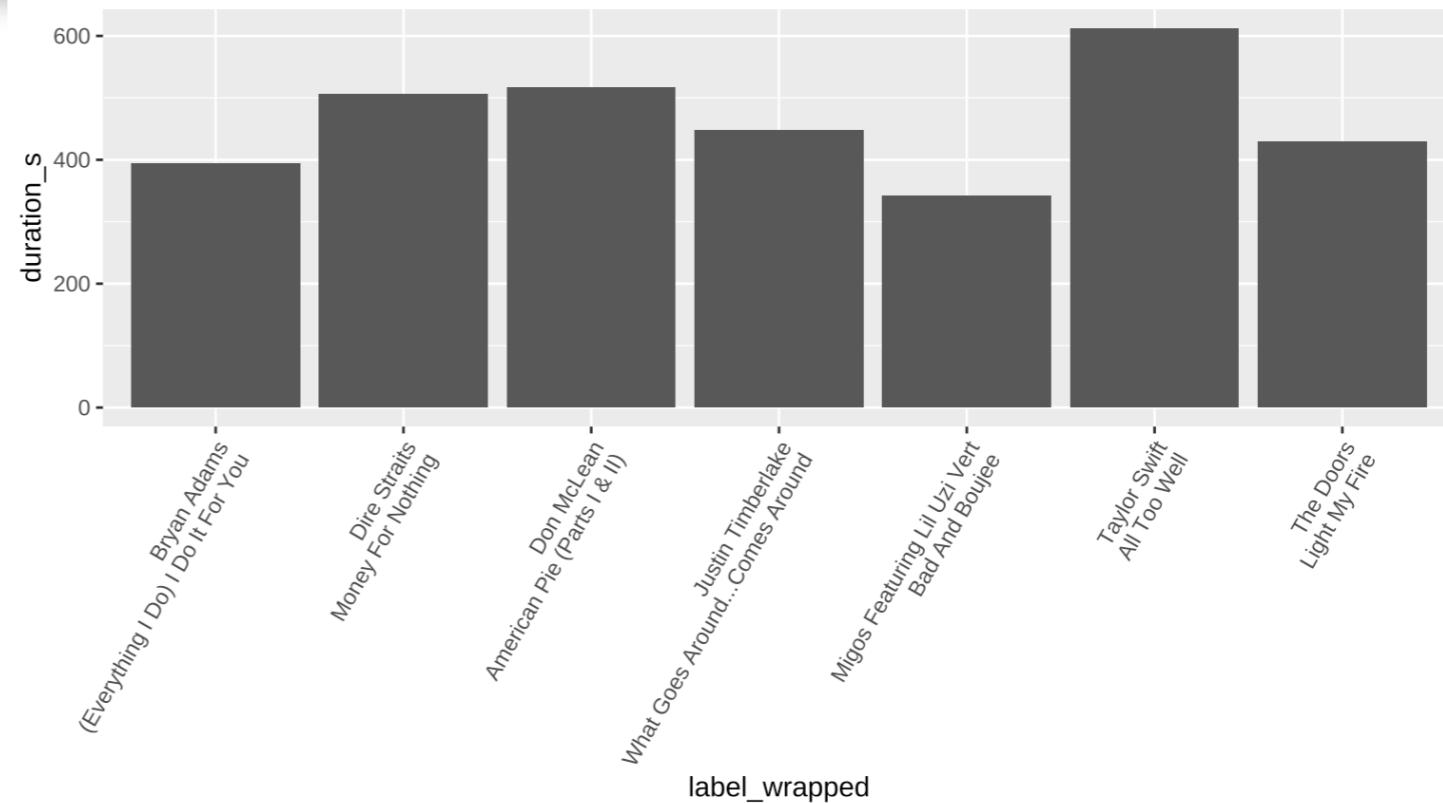
Prepare a dataframe with the longest songs per decade. We'll need these for the labels.

artist_song	artist	song	first_no1_date	decade	duration_s	label	label_wrapped
The Doors - Light My Fire	The Doors	Light My Fire	1967-07-29	1960	429.760	The Doors - Light My Fire	The Doors Light My Fire
Don McLean - American Pie (Parts I & II)	Don McLean	American Pie (Parts I & II)	1972-01-15	1970	516.893	Don McLean - American Pie (Parts I & II)	Don McLean American Pie (Parts I & II)
Dire Straits - Money For Nothing	Dire Straits	Money For Nothing	1985-09-21	1980	506.400	Dire Straits - Money For Nothing	Dire Straits Money For Nothing
Bryan Adams - (Everything I Do) I Do It For You	Bryan Adams	(Everything I Do) I Do It For You	1991-07-27	1990	394.133	Bryan Adams - (Everything I Do) I Do It For You	Bryan Adams (Everything I Do) I Do It For You
Justin Timberlake - What Goes Around...Comes Around	Justin Timberlake	What Goes Around...Comes Around	2007-03-03	2000	448.573	Justin Timberlake - What Goes Around...Comes Around	Justin Timberlake What Goes Around...Comes Around
Migos Featuring Lil Uzi Vert - Bad And Boujee	Migos Featuring Lil Uzi Vert	Bad And Boujee	2017-01-21	2010	343.150	Migos Featuring Lil Uzi Vert - Bad And Boujee	Migos Featuring Lil Uzi Vert Bad And Boujee
Taylor Swift - All Too Well (10 Minute Version) (Taylor's Version)	Taylor Swift	All Too Well (10 Minute Version) (Taylor's Version)	2021-11-27	2020	613.026	Taylor Swift - All Too Well	Taylor Swift All Too Well



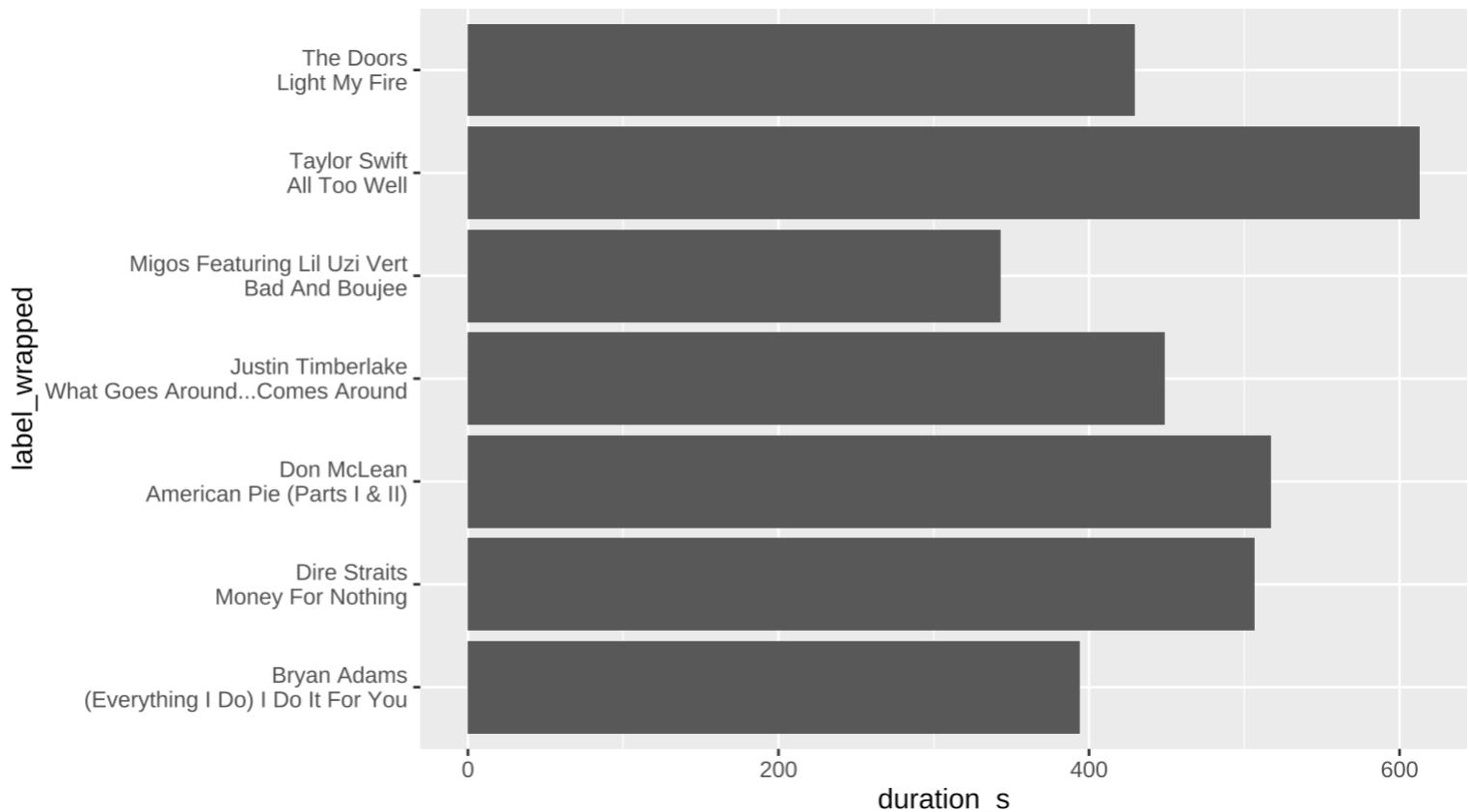
Our first visualization of the data

```
1 ggplot(  
2   data = df_longest_no1_per_decade,  
3   mapping = aes(x = label, y = duration_s)) +  
4   geom_col()
```



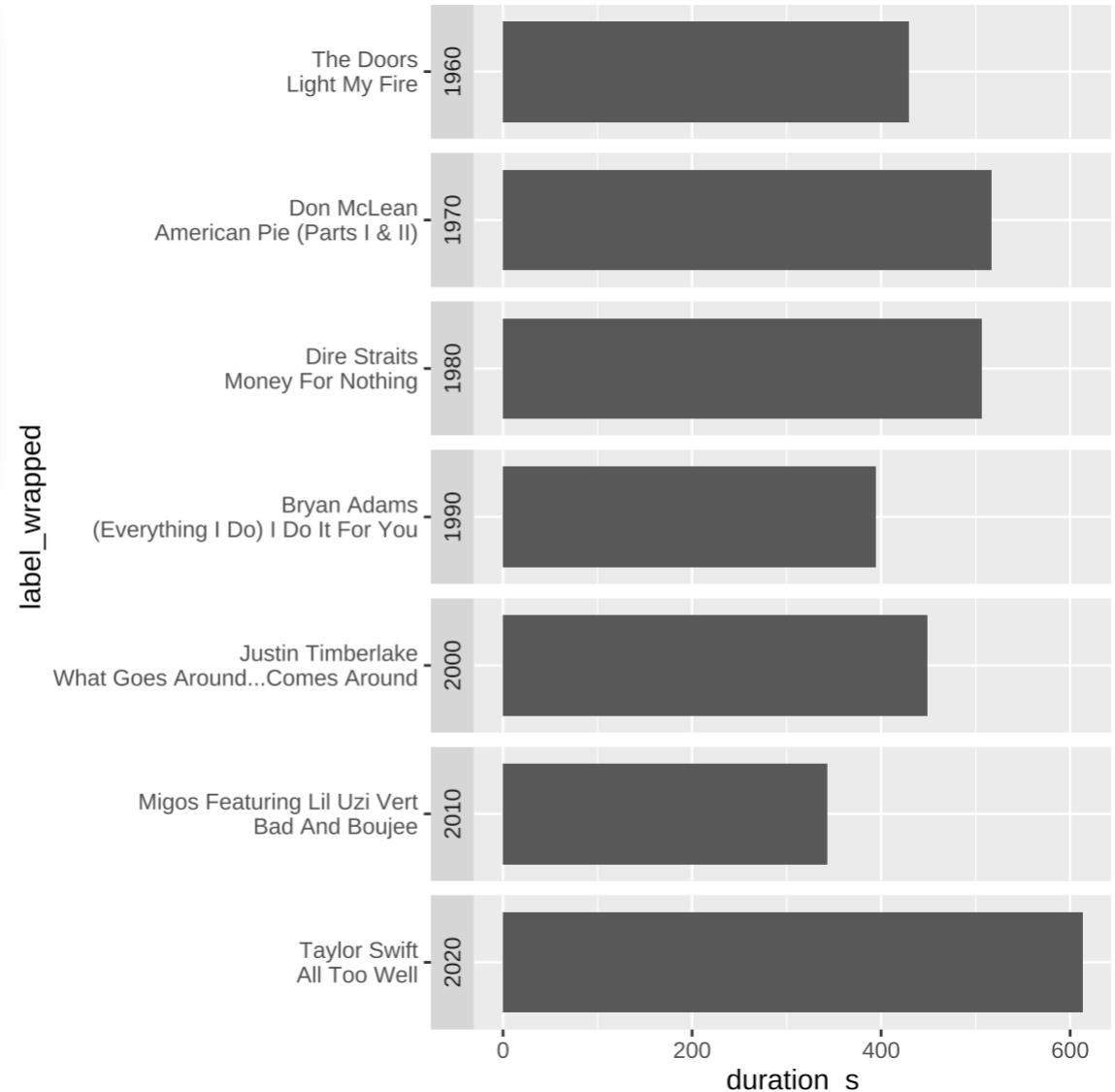
Dataviz 101 - Rotate that damn plot!

```
1 ggplot(  
2   data = df_longest_nol_per_decade,  
3   mapping = aes(x = duration_s,  
4                 y = label_wrapped)) +  
5   geom_col()
```



We haven't mentioned the decades so far

```
1 p <- ggplot(  
2   data = df_longest_nol_per_decade,  
3   mapping = aes(x = duration_s,  
4                 y = label_wrapped)) +  
5   geom_col() +  
6   facet_wrap(  
7     vars(decade),  
8     ncol = 1, scales = "free_y",  
9     strip.position = "left")
```



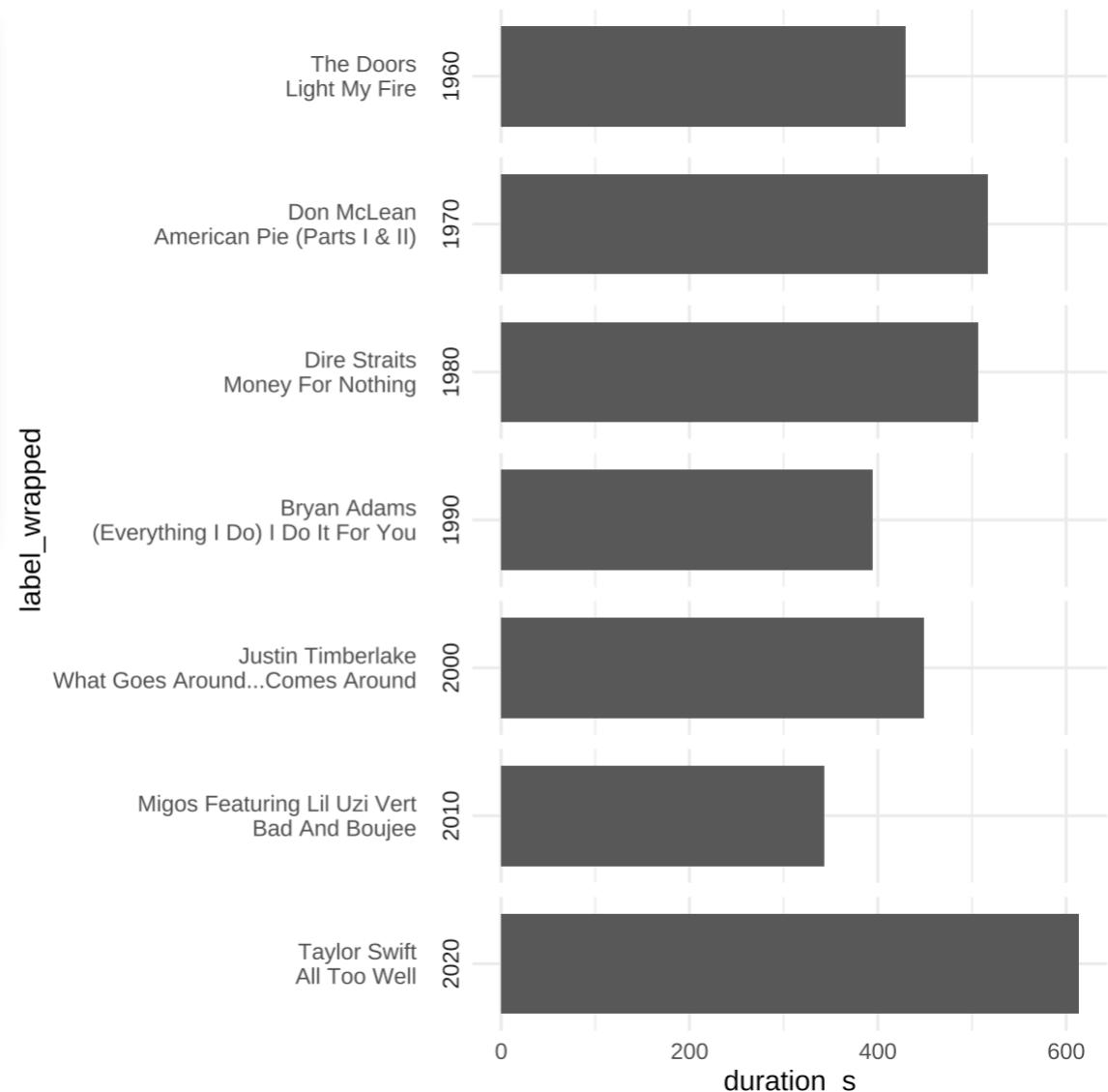
Apply a theme

```
1 p <- ggplot(  
2   data = df_longest_no1_per_decade,  
3   mapping = aes(x = duration_s,  
4                   y = label_wrapped)) +  
5   geom_col() +  
6   facet_wrap(  
7     vars(decade),  
8     ncol = 1, scales = "free_y",  
9     strip.position = "left") +  
10  theme_minimal()
```

{ggplot2} contains a couple of different themes out of the box that you can apply to your plot, e.g.:

- `theme_bw()`
- `theme_light()`

+ additional packages available



Switch the song titles to the right

```
1 p <- ggplot(  
2   data = df_longest_nol_per_decade,  
3   mapping = aes(x = duration_s,  
4                 y = label_wrapped)) +  
5   geom_col() +  
6   scale_y_discrete(position = "right") +  
7   facet_wrap(  
8     vars(decade),  
9     ncol = 1, scales = "free_y",  
10    strip.position = "left") +  
11   theme_minimal()
```

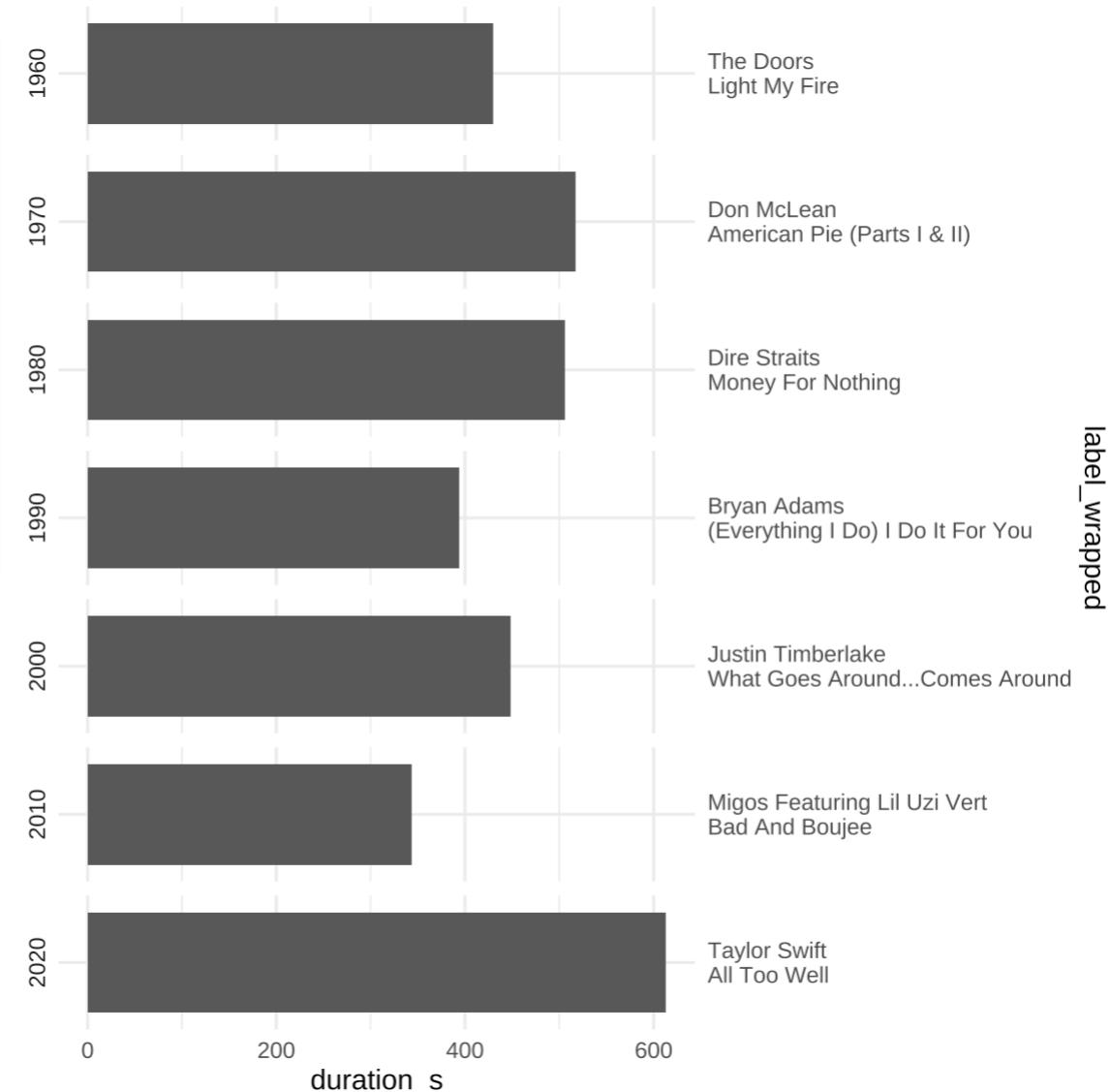
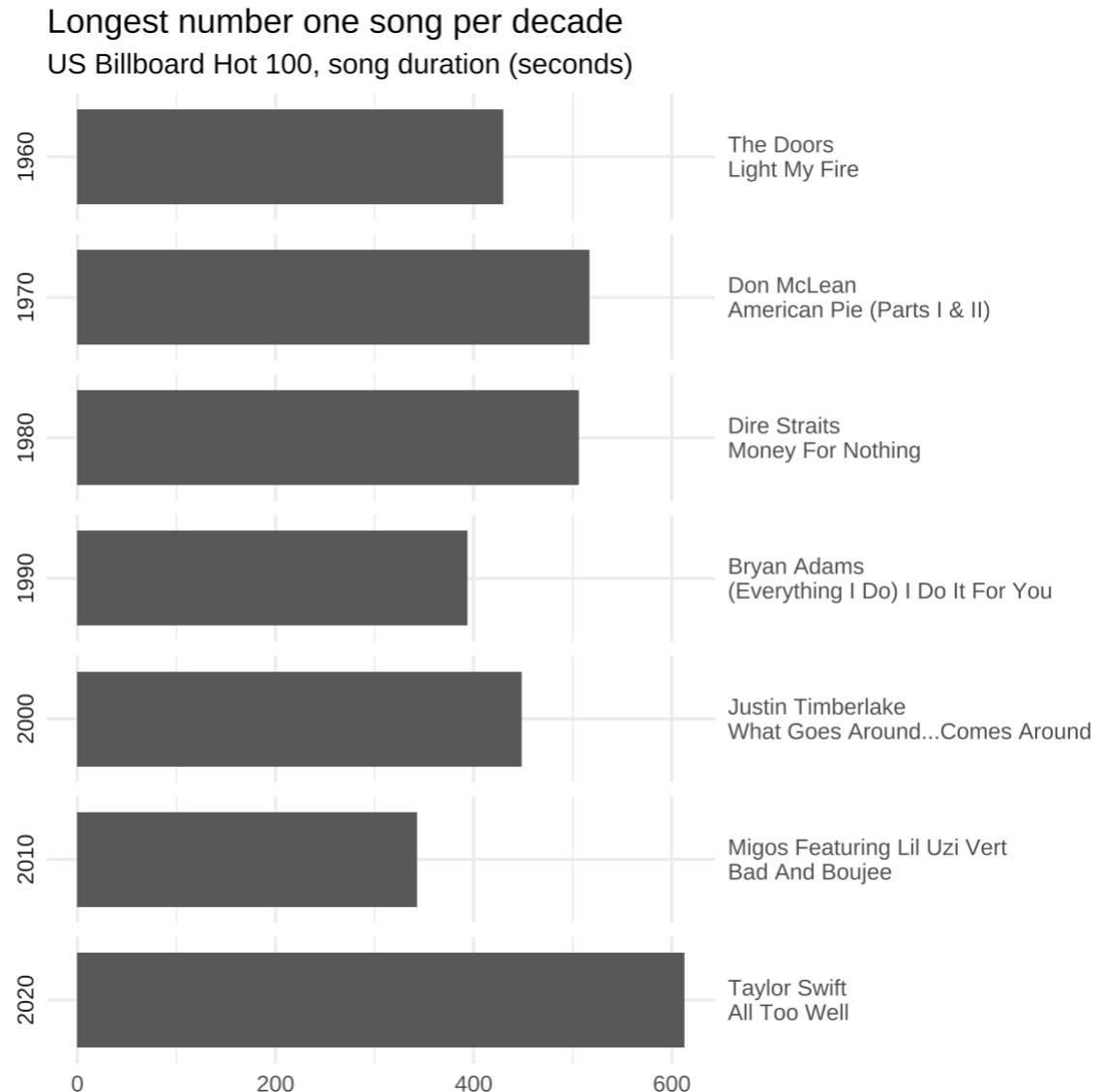


Chart title and labels

```
1 p <- ggplot(  
2   data = df_longest_no1_per_decade,  
3   mapping = aes(x = duration_s,  
4                   y = label_wrapped)) +  
5   geom_col() +  
6   scale_y_discrete(position = "right") +  
7   facet_wrap(  
8     vars(decade),  
9     ncol = 1, scales = "free_y",  
10    strip.position = "left") +  
11   labs(  
12     title = "Longest number one song per decade",  
13     subtitle = "US Billboard Hot 100, song duration (seconds)",  
14     x = NULL,  
15     y = NULL  
16   ) +  
17   theme_minimal()
```

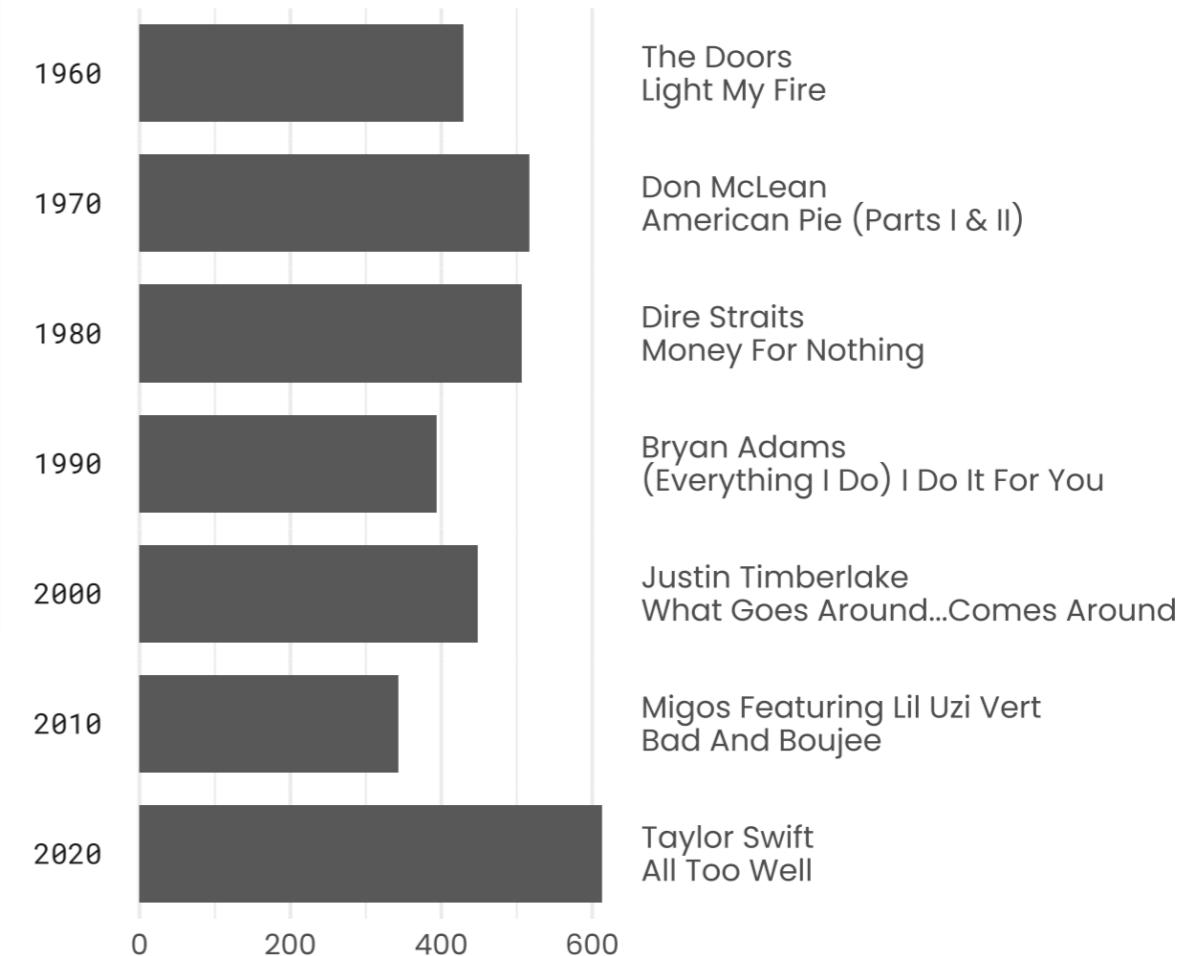


More theming

```
1 p <- p +  
2   theme_minimal(  
3     base_family = "Poppins", base_size = 14)  
4 theme(  
5   strip.text.y.left = element_text(  
6     angle = 0, family = "Roboto Mono"),  
7   plot.title = element_text(face = "bold"),  
8   plot.subtitle = element_markdown(),  
9   plot.title.position = "plot",  
10  panel.grid.major.y = element_blank(),  
11  panel.grid.minor.y = element_blank(),  
12  panel.spacing.y = unit(0, "mm"),  
13  axis.text.y.right = element_text(size = 1  
14 )
```

Longest number one song per decade

US Billboard Hot 100, song duration (seconds)



What if we wanted to show the distribution?

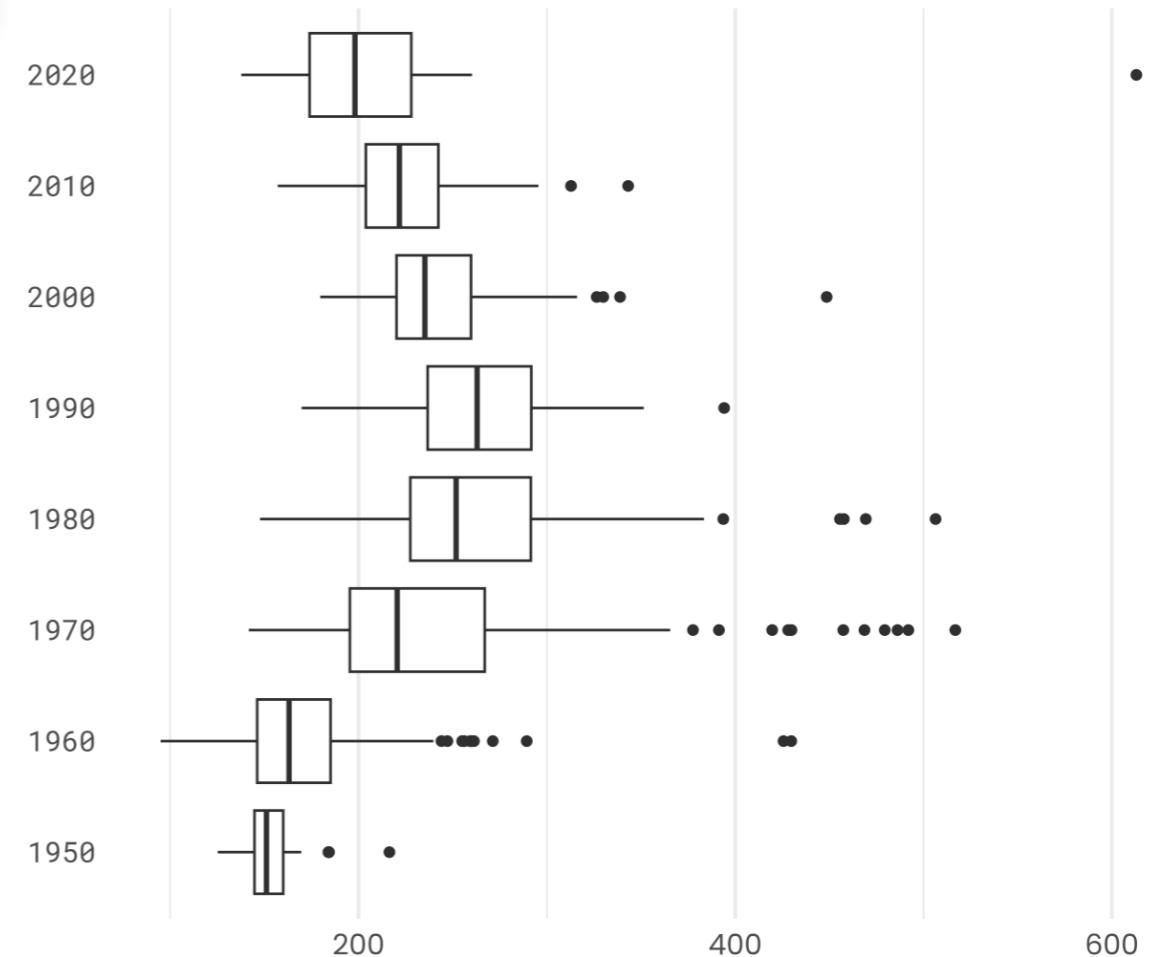


Boxplots?

```
1 p <- p_base +  
2   geom_boxplot()
```

Length of number 1 songs per decade

US Billboard Hot 100, song duration in seconds



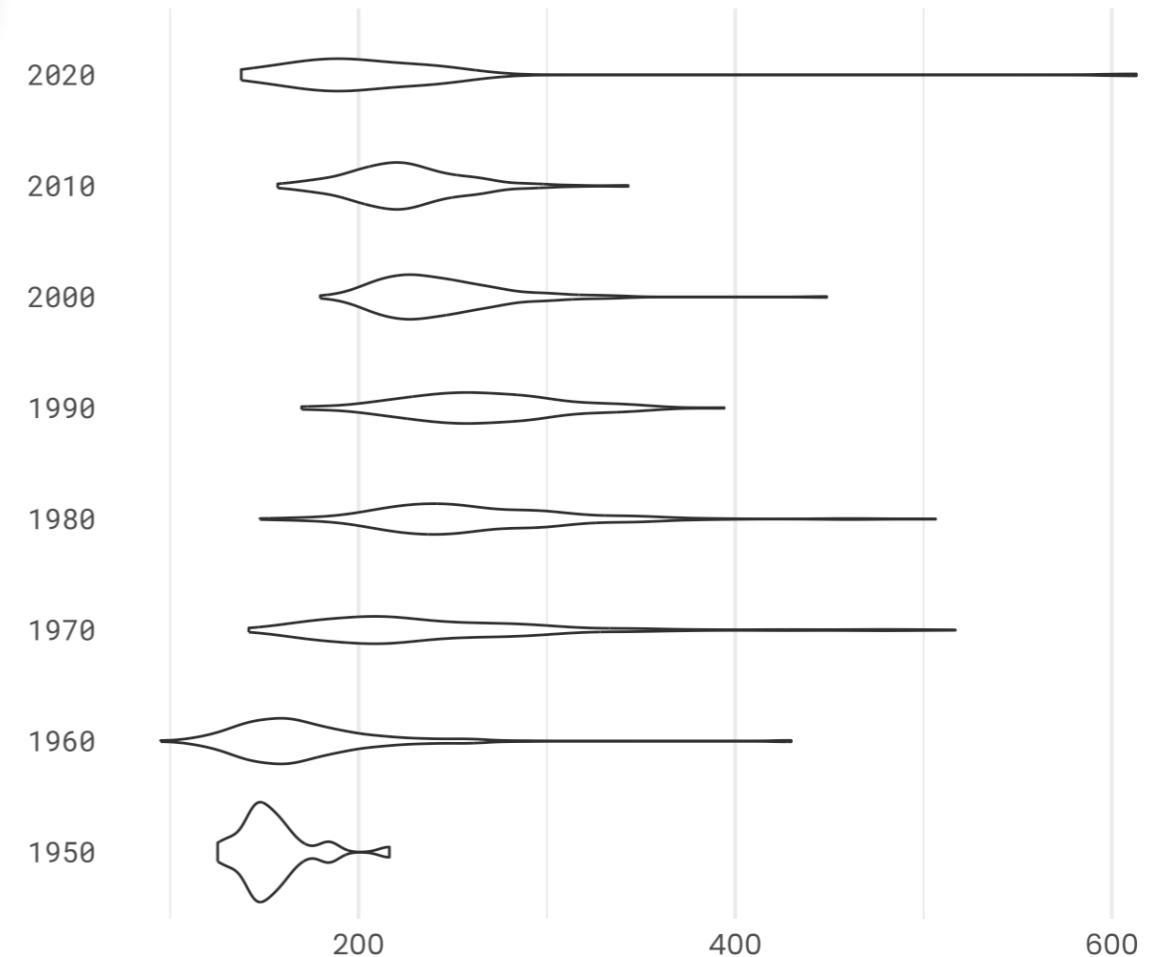
Music? Violins!



```
1 p <- p_base +  
2   geom_violin()
```

Length of number 1 songs per decade

US Billboard Hot 100, song duration in seconds



Density-interval combo (“halfeye”)?

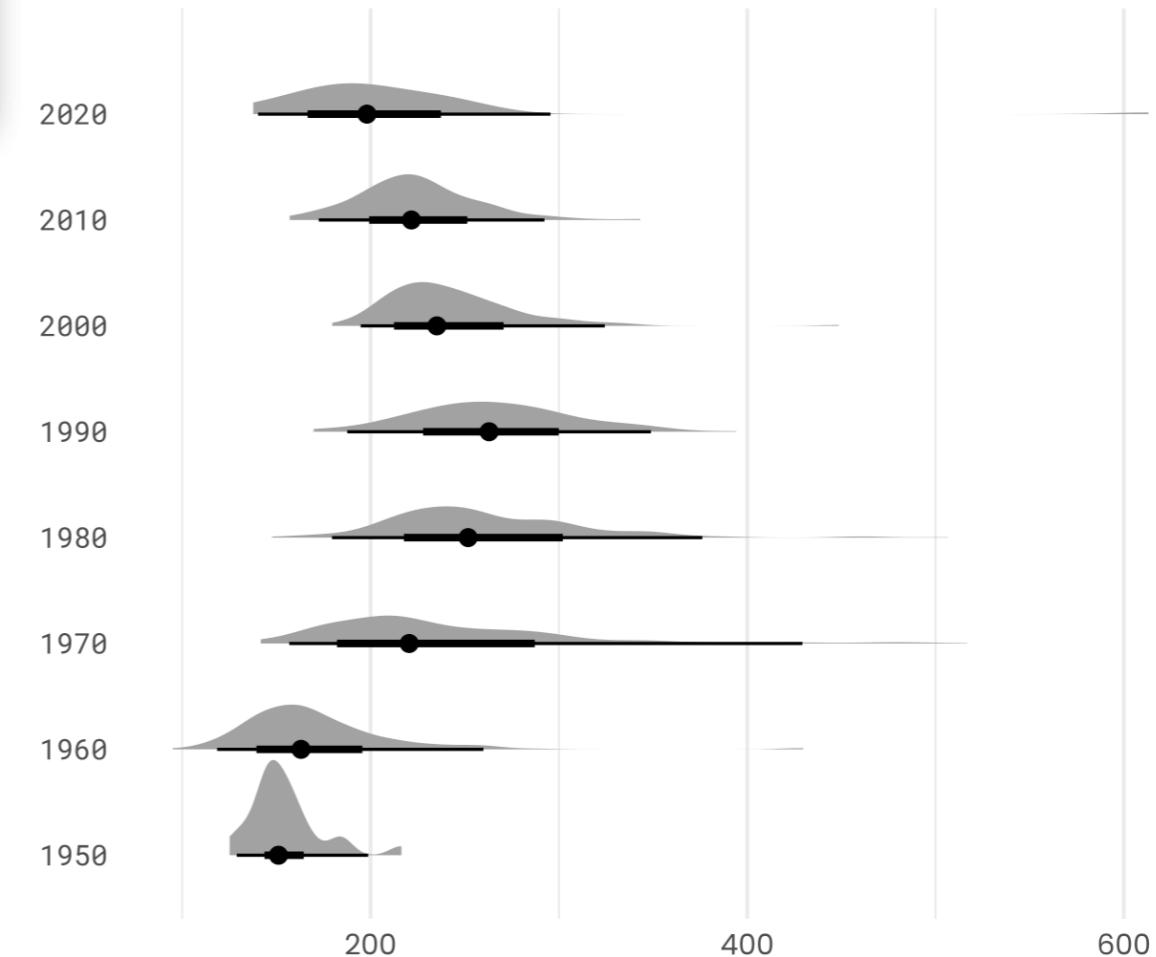
```
1 library(ggdist)
2
3 p <- p_base +
4   stat_halfeye()
```

{ggdist} provides functions which extended {ggplot2} to visualize...

- distributions
- uncertainty

Length of number 1 songs per decade

US Billboard Hot 100, song duration in seconds



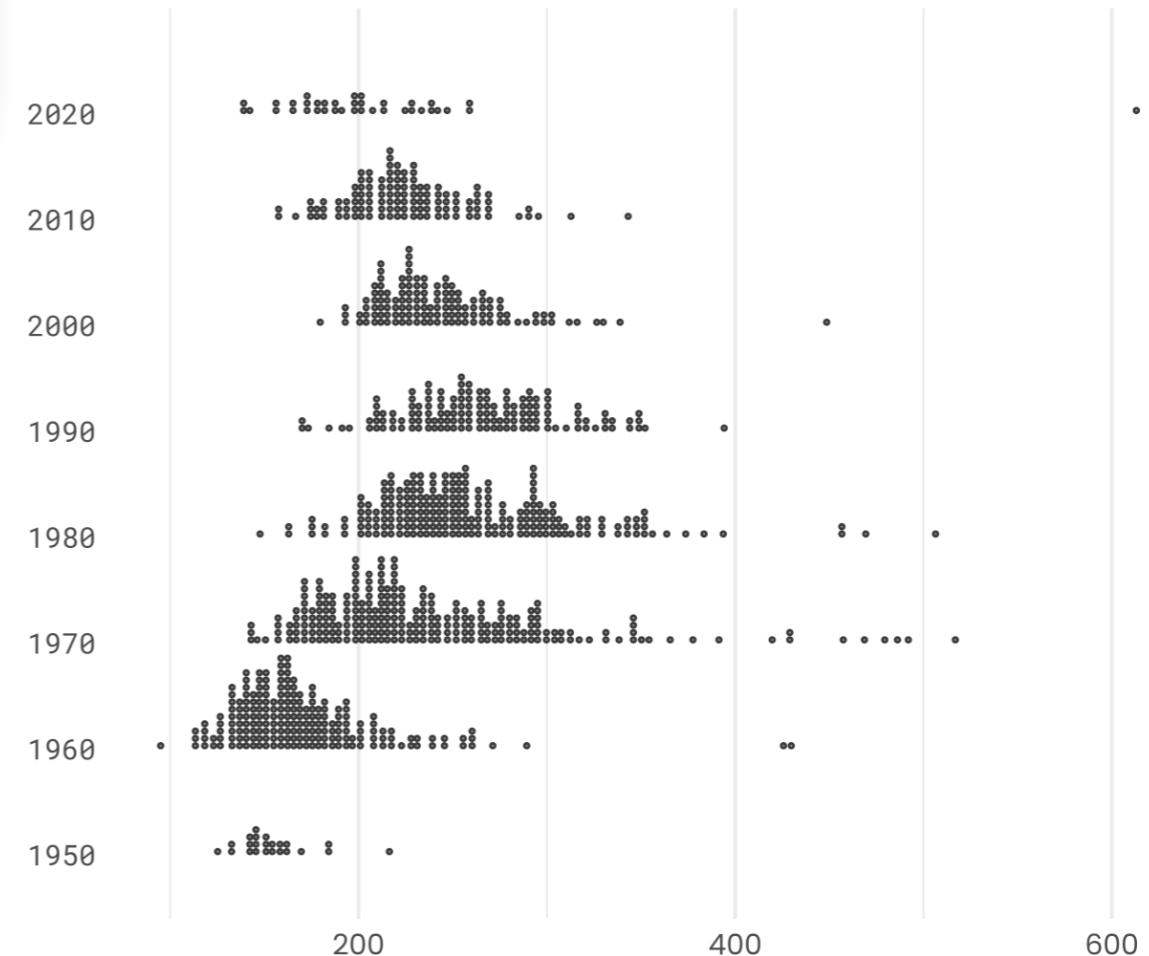
Show every single data point?

```
1 p <- p_base +  
2   stat_dots(  
3     dotsize = 1.2, color = "grey24"  
4   )
```

`stat_dots()` is also from the `{ggdist}` package

Length of number 1 songs per decade

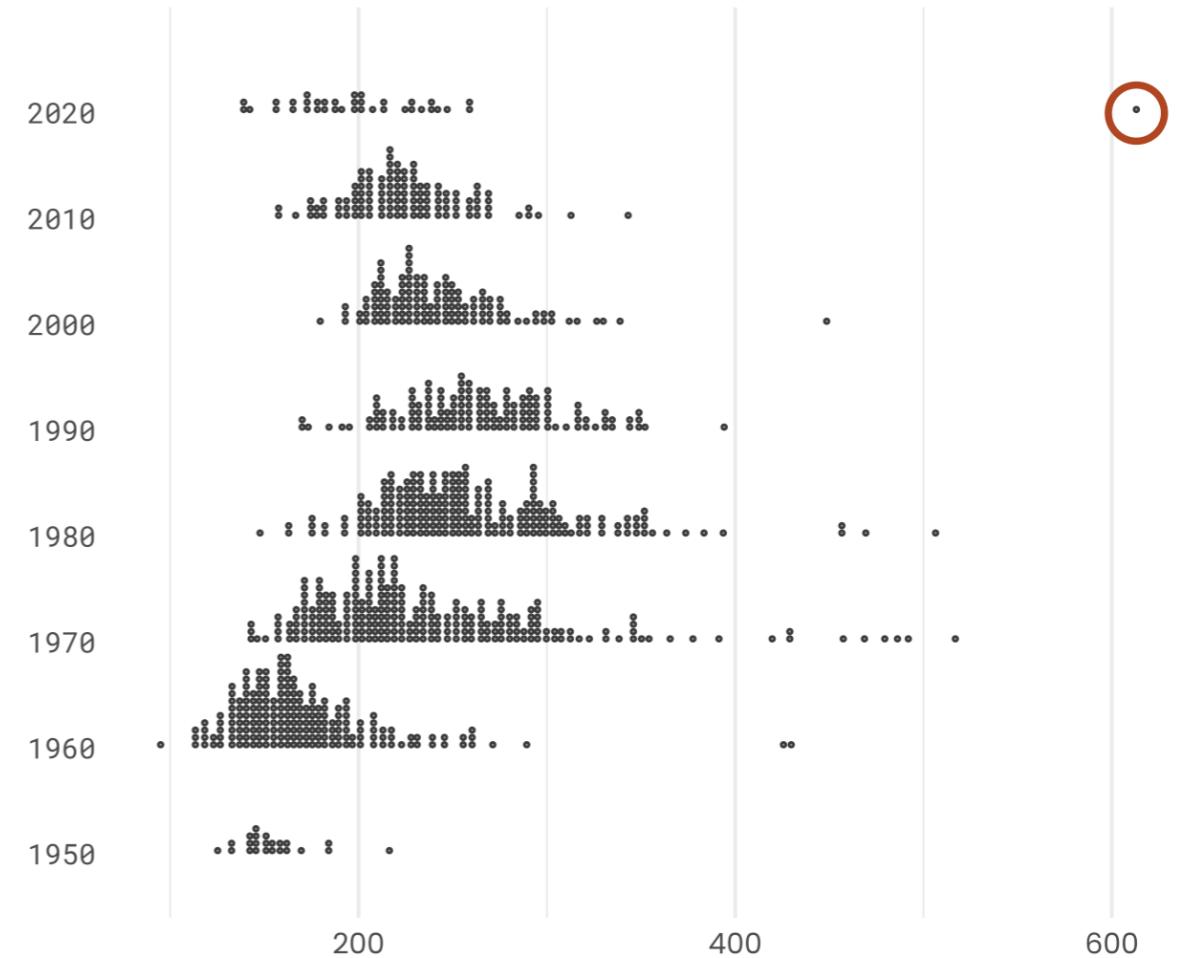
US Billboard Hot 100, song duration in seconds



Where's Waldo Taylor?

Length of number 1 songs per decade

US Billboard Hot 100, song duration in seconds

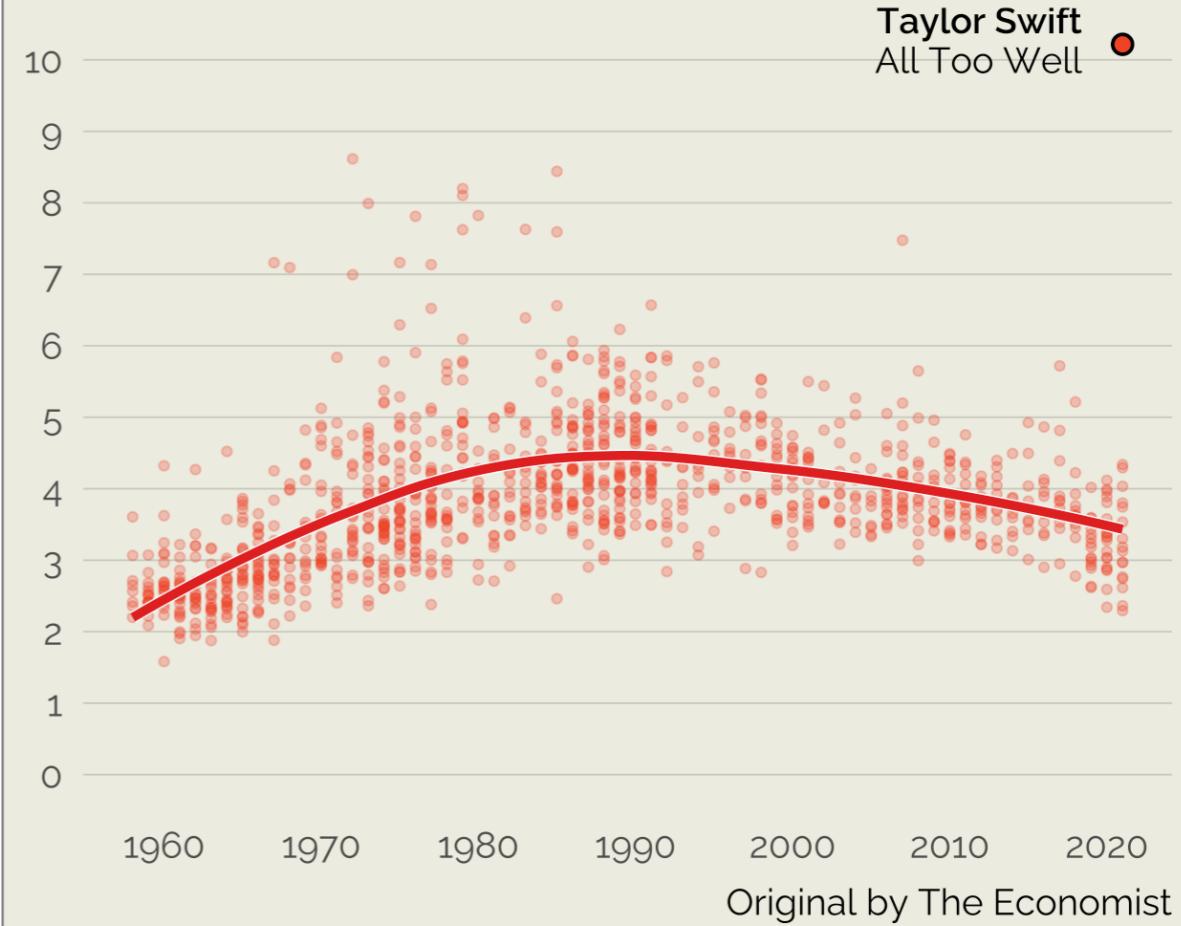


Like the Economist (2025)

```
1 p <- tracks_df |>
2   mutate(
3     year = year(first_no1_date),
4     duration_min = duration_s / 60) |>
5   ggplot(
6     mapping = aes(
7       x = year, y = duration_min)
8   ) +
9   geom_point(
10    color = "#F04629",
11    alpha = 0.3) +
12   geom_smooth(
13     method = "loess",
14     color = "white", linewidth = 2.2,
15     se = FALSE) +
16   geom_smooth(
17     method = "loess",
18     color = "#DF2124", linewidth = 1.7,
19     se = FALSE) +
```

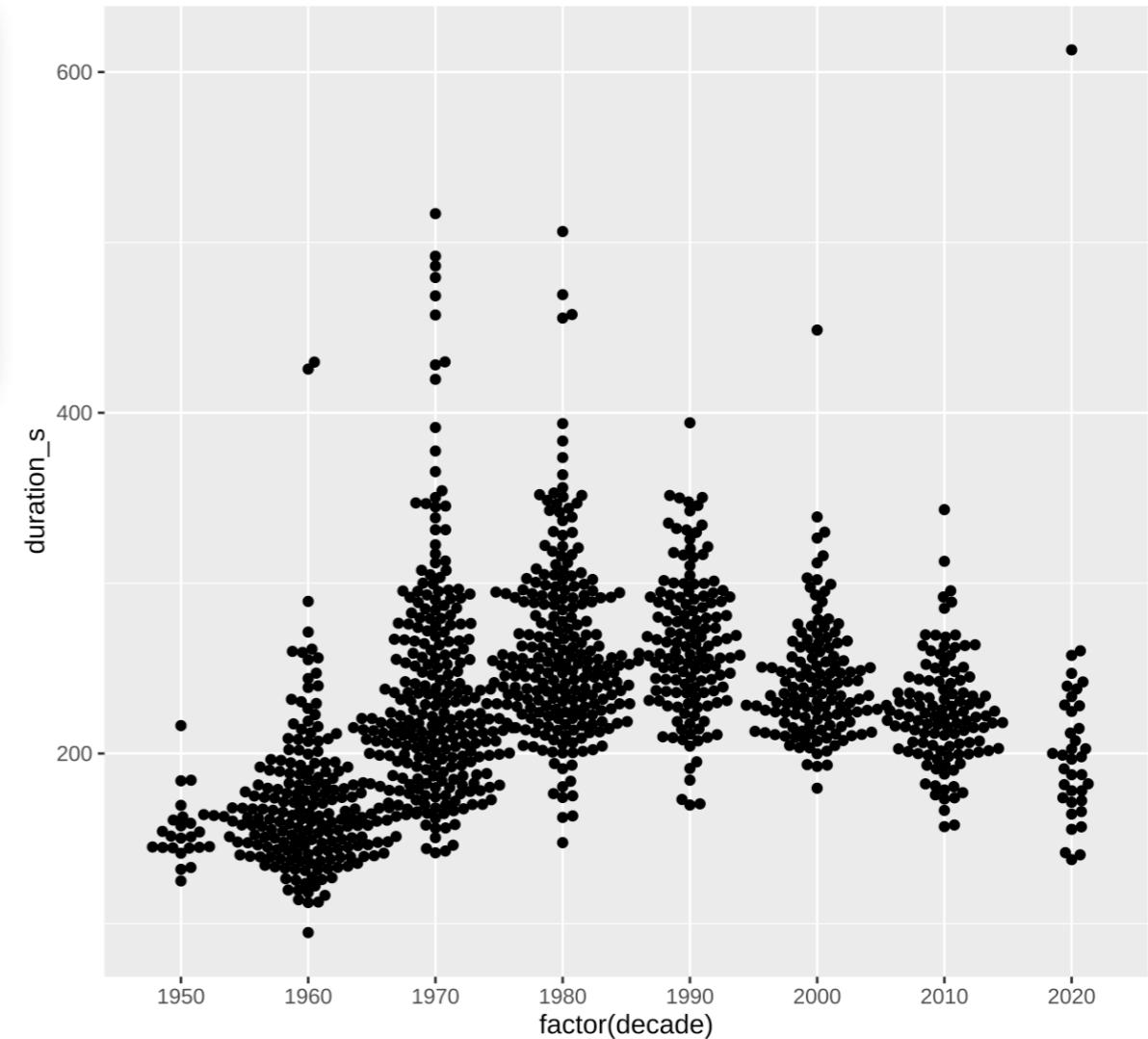
Turning back time

US Billboard Hot 100, Length in minutes



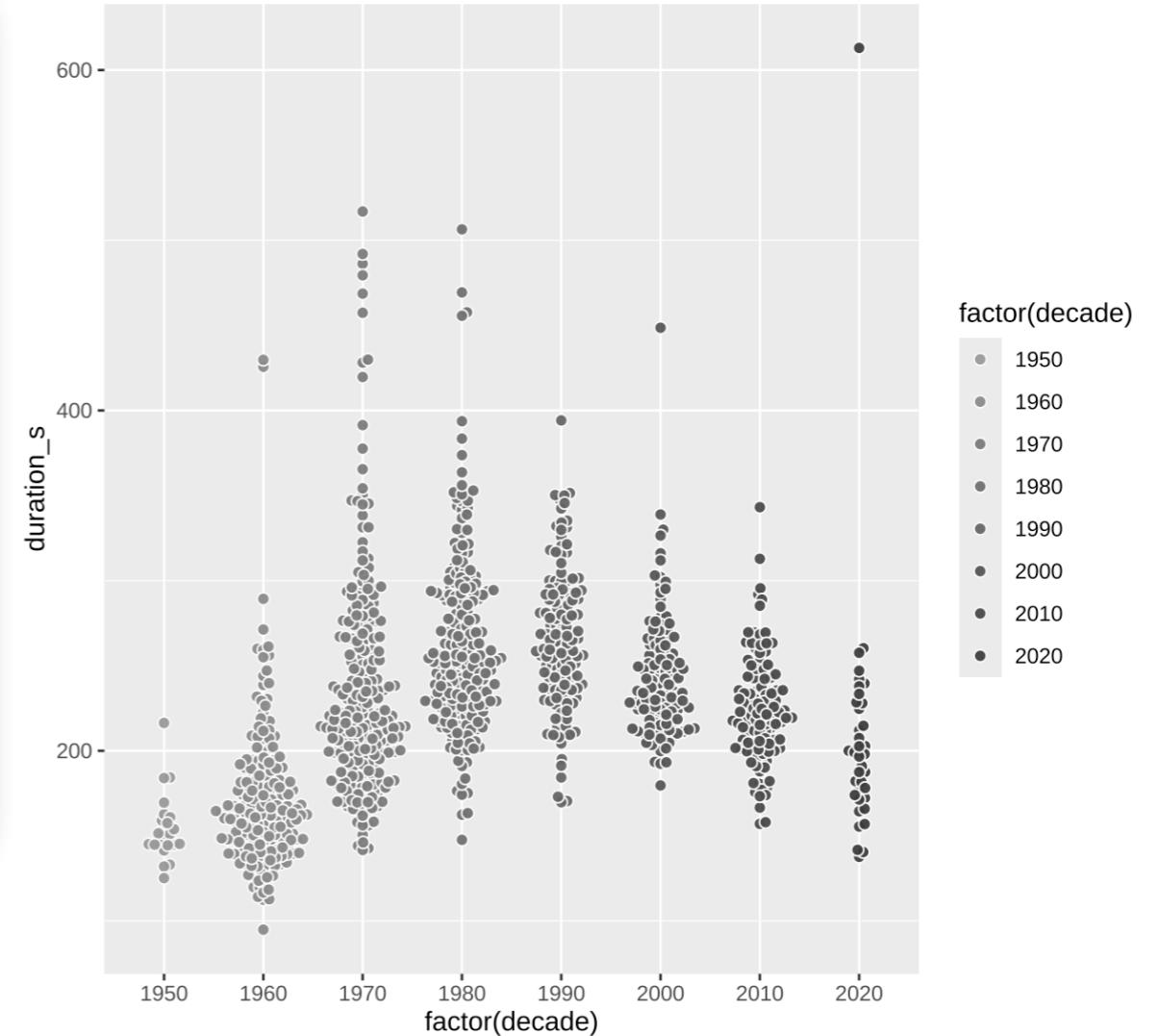
Let's start creating the plot

```
1 library(ggbeeswarm)  
2  
3 p <- ggplot(  
4   data = tracks_df,  
5   aes(x = factor(decade),  
6        y = duration_s)) +  
7   geom_beeswarm()
```



Style the dots

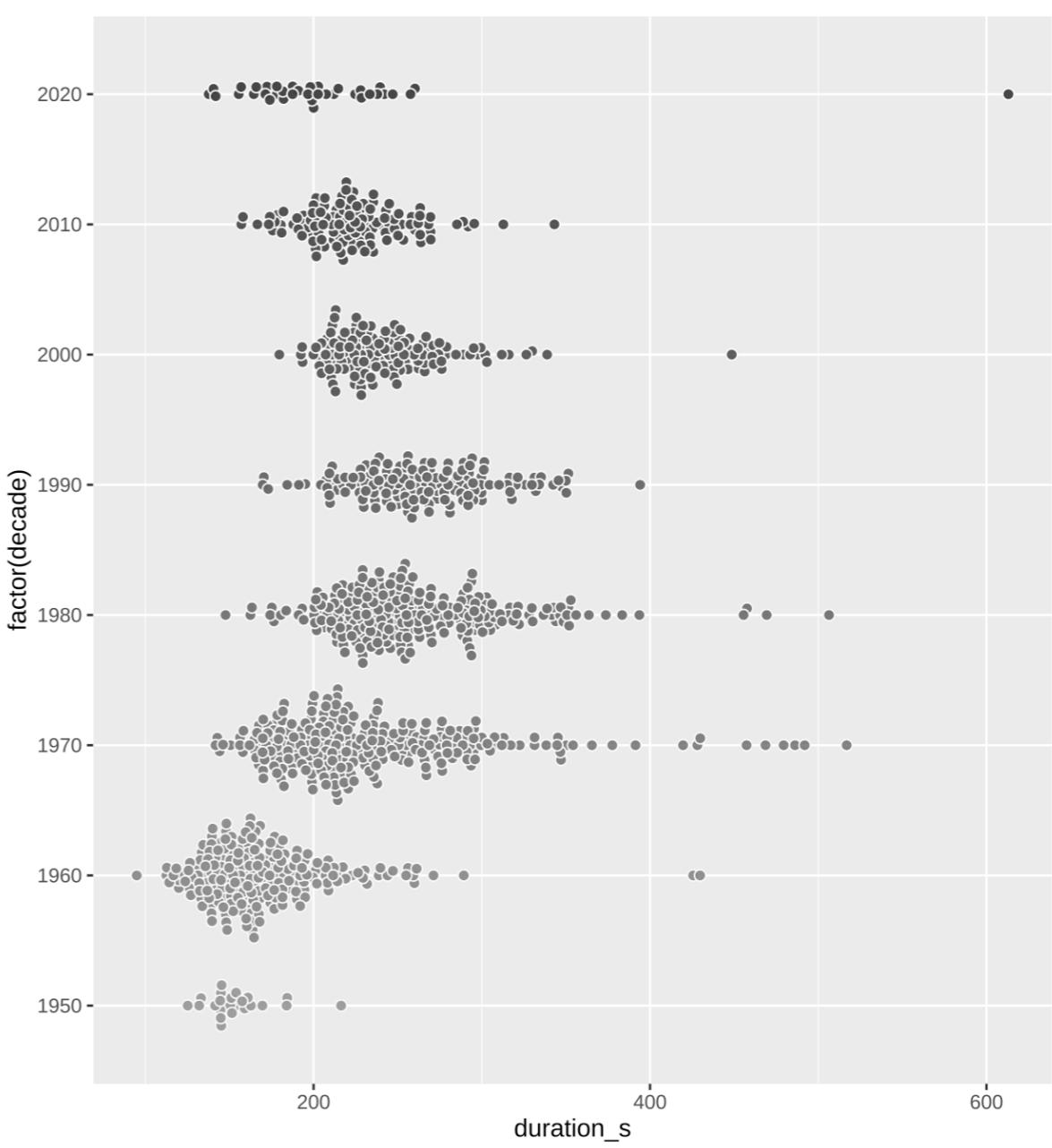
```
1 greyscale_palette <- c(  
2   "grey60", "grey55", "grey50", "grey45",  
3   "grey40", "grey35", "grey30", "grey25"  
4 )  
5  
6 p <- ggplot(  
7   data = tracks_df,  
8   aes(x = factor(decade),  
9     y = duration_s)) +  
10  geom_beeswarm(  
11    aes(fill = factor(decade)),  
12    cex = 0.75, shape = 21,  
13    col = "white", alpha = 0.95,  
14    size = 2) +  
15  scale_fill_manual(  
16    values = greyscale_palette)
```



Rotate the plot and remove the legend



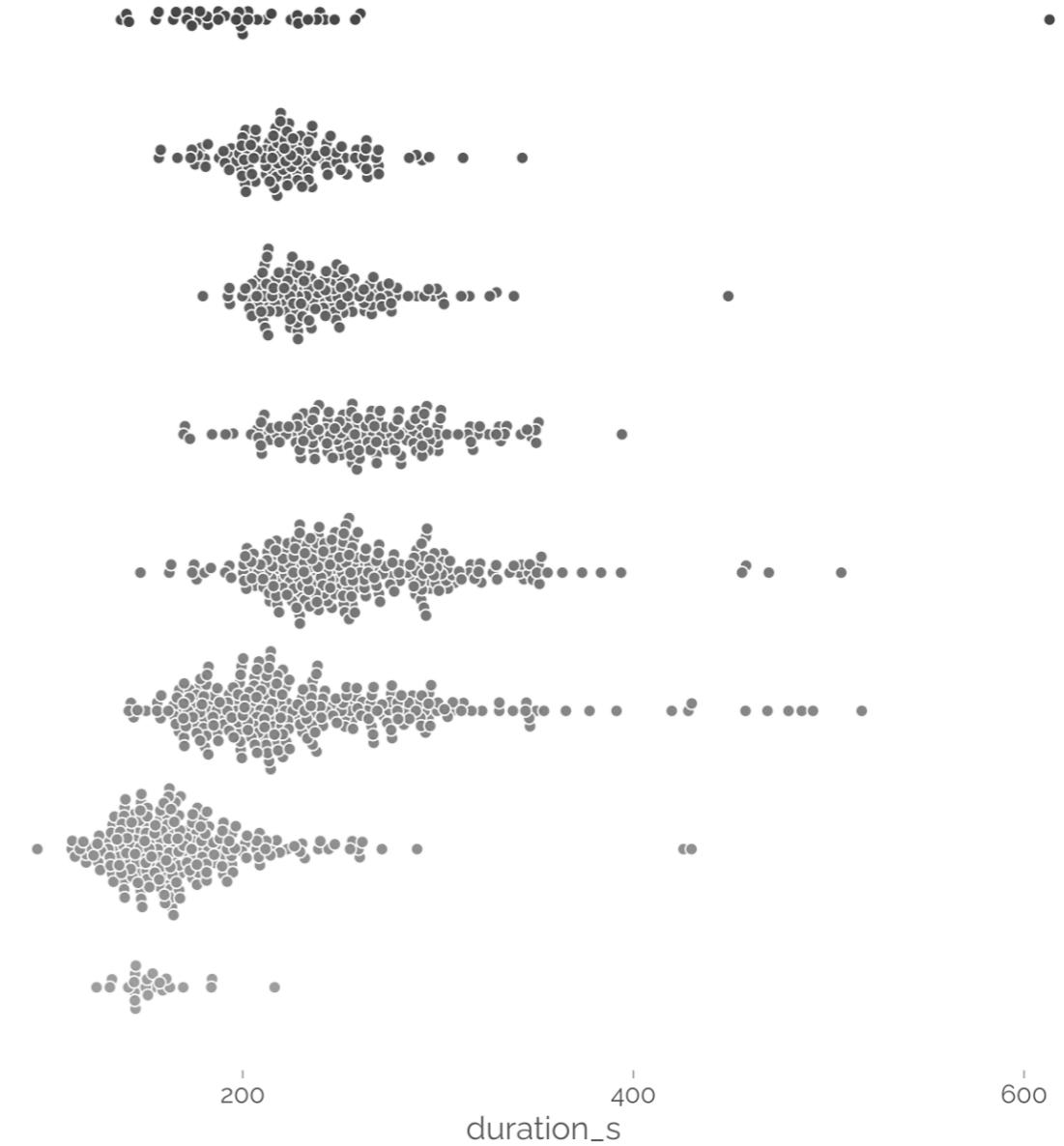
```
1 p <- p +  
2 coord_flip() +  
3 guides(fill = "none")
```



Add the theme



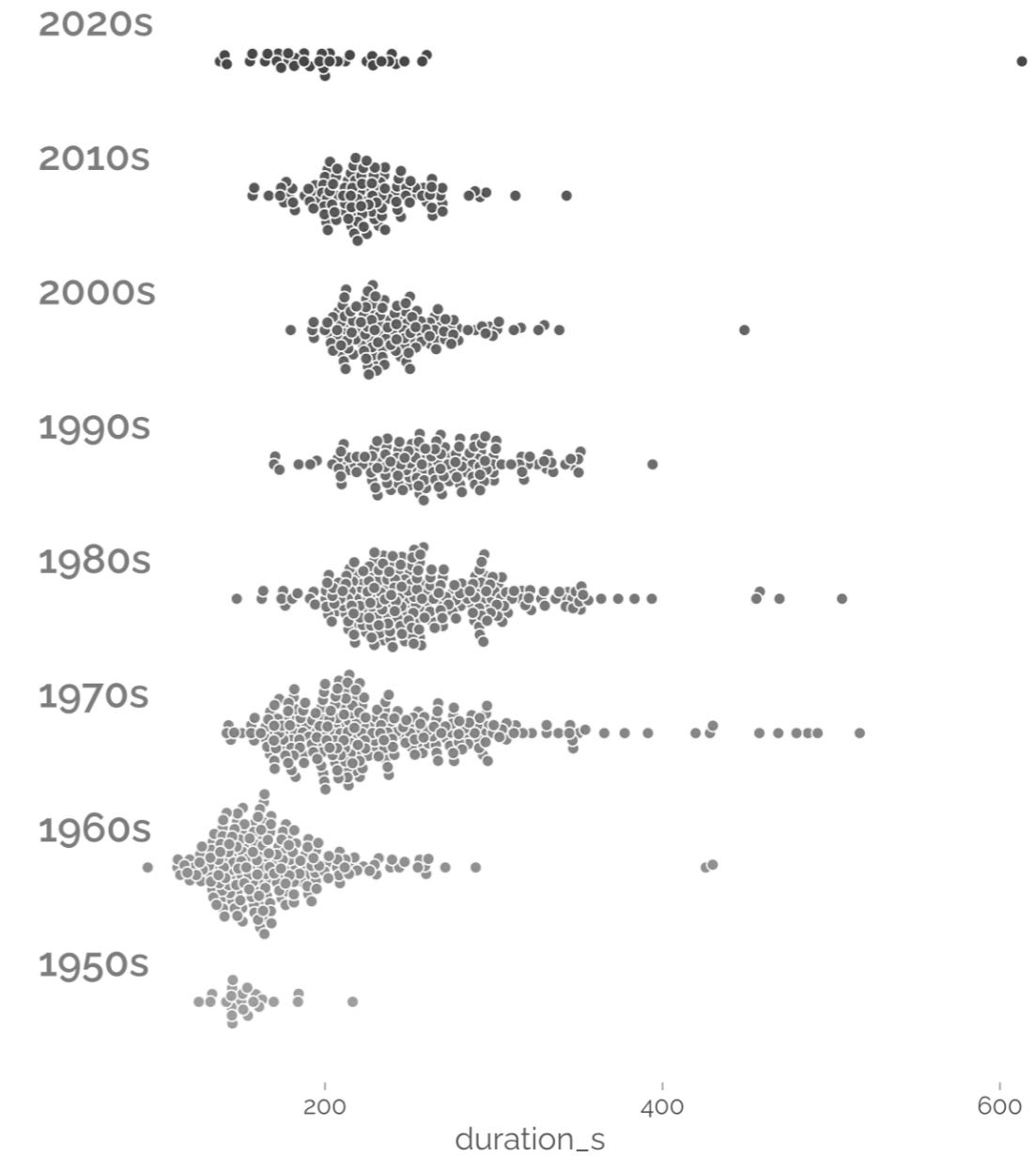
```
1 p <- p +  
2   theme_minimal(  
3     base_family = "Raleway",  
4     base_size = 13) +  
5   theme(  
6     plot.background = element_rect(  
7       color = "white", fill = "white"),  
8     panel.grid = element_blank(),  
9     axis.title.y = element_blank(),  
10    axis.text.y = element_blank(),  
11    axis.ticks.x = element_line(  
12      color = "grey60", linewidth = 0.3),  
13    text = element_text(  
14      color = "grey35"),  
15    plot.title = element_markdown(  
16      color = "black",  
17      family = "Raleway SemiBold", size = 18),  
18    plot.subtitle = element_textbox(  
19      margin = margin(t = 6, b = 6),
```



Add annotations for the decades



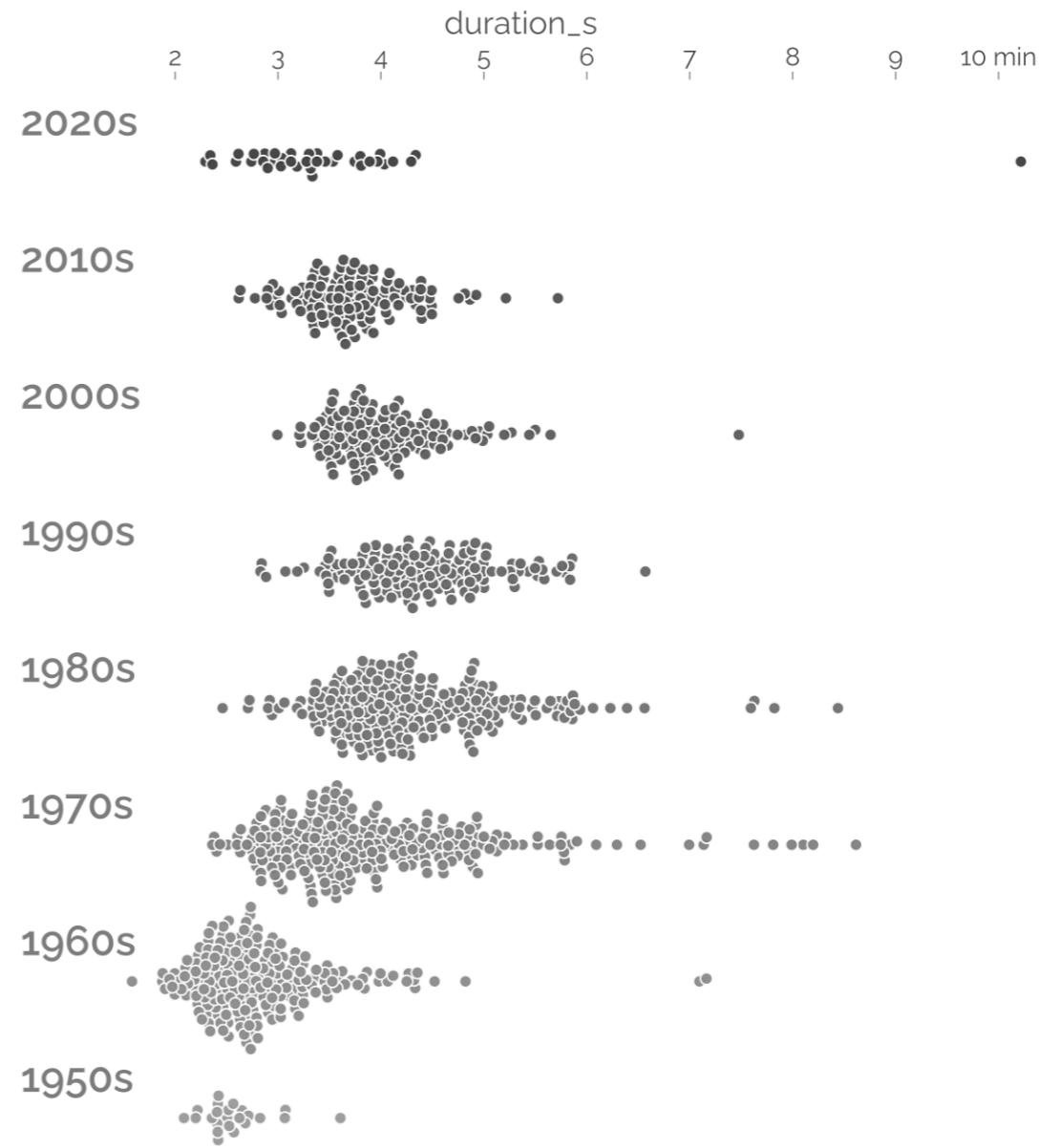
```
1 p <- p +  
2 # Labels for decades  
3 annotate(  
4   "text",  
5   x = seq_along(seq(1950, 2020, 10)) + 0.3,  
6   y = 30,  
7   label = paste0(seq(1950, 2020, 10), "s"),  
8   family = "Raleway SemiBold",  
9   color = "grey50", size = 6,  
10  hjust = 0  
11 )
```



Format the x-axis



```
1 p <- p +  
2   scale_y_continuous(  
3     position = "right",  
4     breaks = seq(120, 600, 60),  
5     labels = c(  
6       seq(120, 540, 60) / 60,  
7       "10 min")  
8   )
```



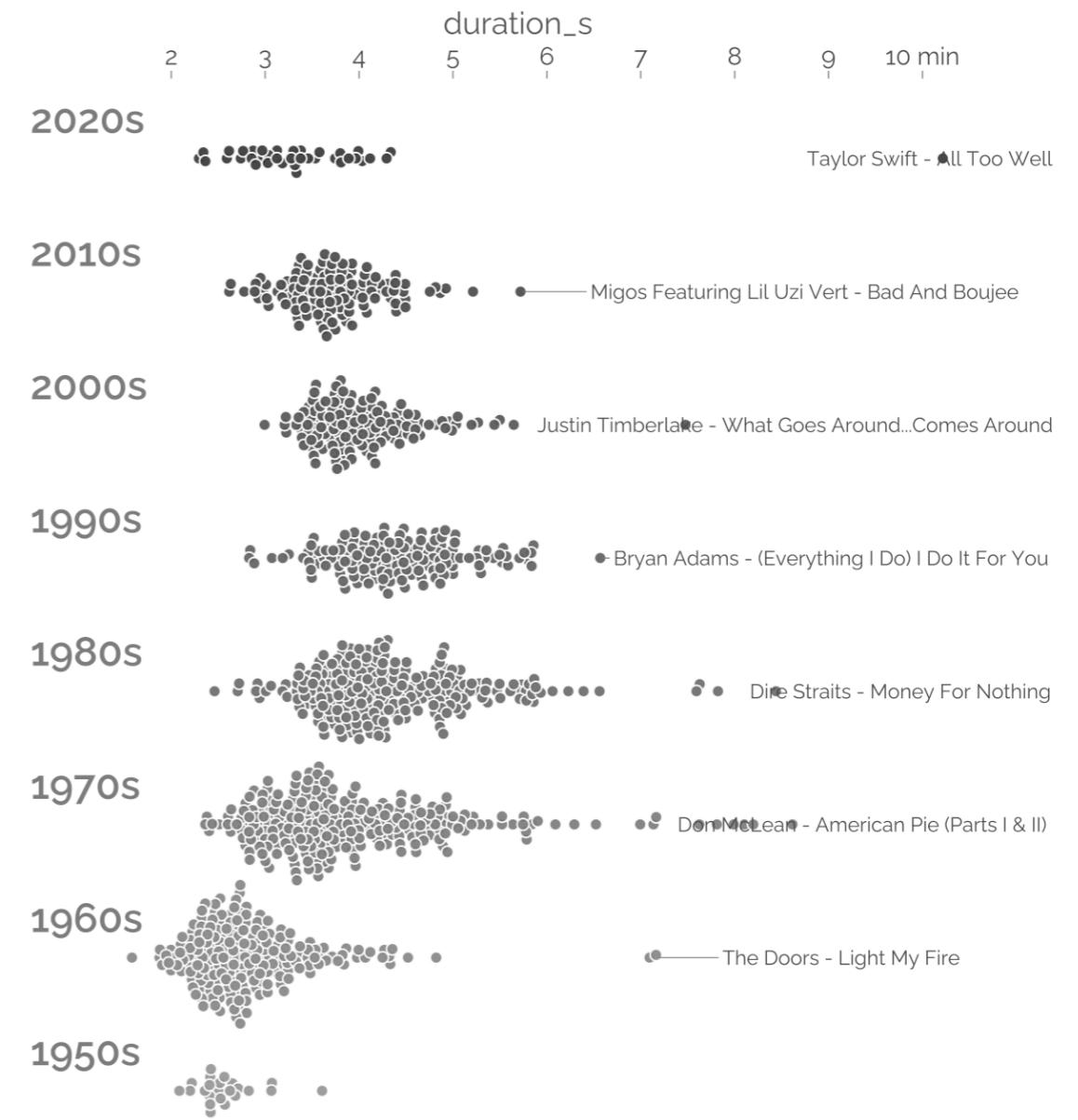
Add annotations for the longest songs



```

1 p <- p +
2   ggrepel::geom_text_repel(
3     data = df_longest_no1_per_decade,
4     aes(label = label),
5     size = 3, family = "Raleway",
6     color = "grey30", lineheight = 0.8,
7     min.segment.length = unit(0.1, "mm"),
8     segment.size = 0.2,
9     segment.color = "grey50",
10    direction = "y", hjust = 0,
11    nudge_y = 40
12 )

```



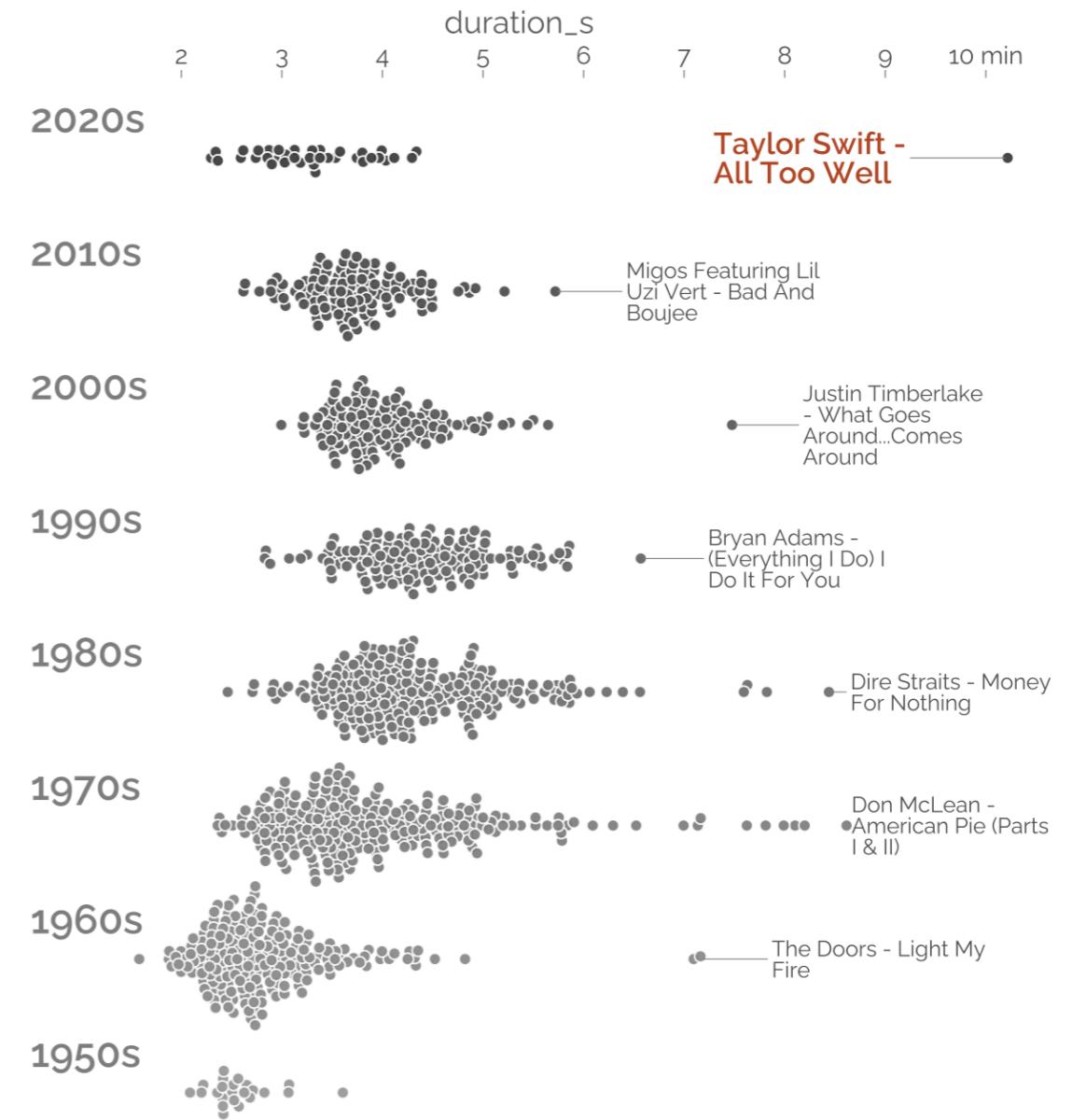
We can do better



```

1 p <- p +
2   ggrepel::geom_text_repel(
3     data = filter(
4       df_longest_no1_per_decade,
5       id == track_id),
6     aes(label = str_wrap(label, 16)),
7     size = 4.5, family = "Raleway SemiBold",
8     color = highlight_color, lineheight = 0.8
9     min.segment.length = unit(0.1, "mm"),
10    segment.size = 0.2,
11    segment.color = "grey50",
12    direction = "y", hjust = 0,
13    nudge_y = -180
14  ) +
15   ggrepel::geom_text_repel(
16     data = filter(
17       df_longest_no1_per_decade,
18       id != track_id),
19     aes(label = str_wrap(label, 20)),

```



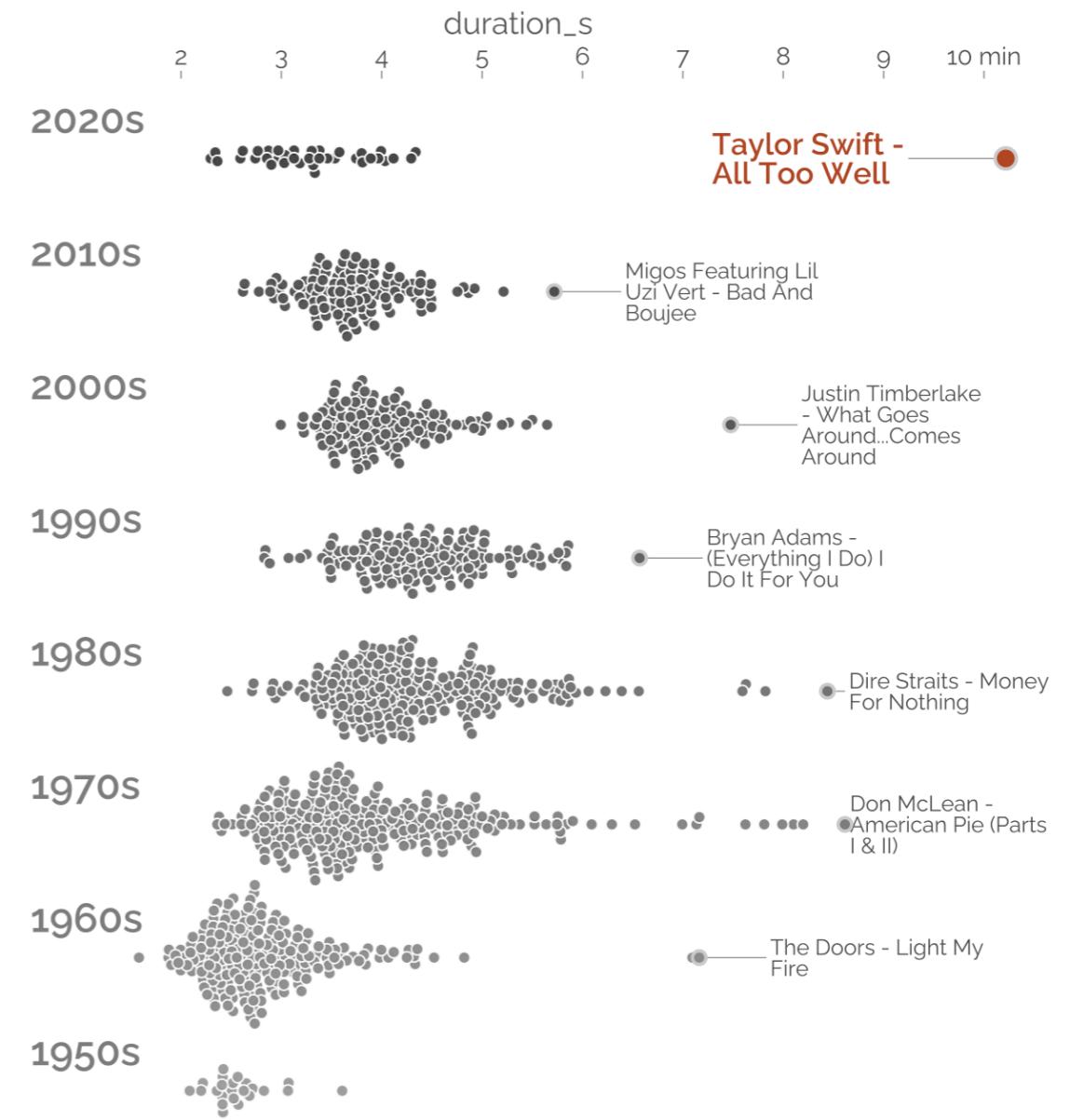
Highlight the longest songs



```

1 p <- p +
2   geom_point(
3     data = filter(
4       df_longest_no1_per_decade,
5       id != track_id),
6     aes(fill = factor(decade)),
7     shape = 21, stroke = 1,
8     color = "grey80", size = 2,
9     show.legend = FALSE) +
10  geom_point(
11    data = filter(
12      df_longest_no1_per_decade,
13      id == track_id),
14    shape = 21, stroke = 1,
15    color = "grey80", size = 3.5,
16    fill = highlight_color,
17    show.legend = FALSE)

```



Finally, add the titles



```

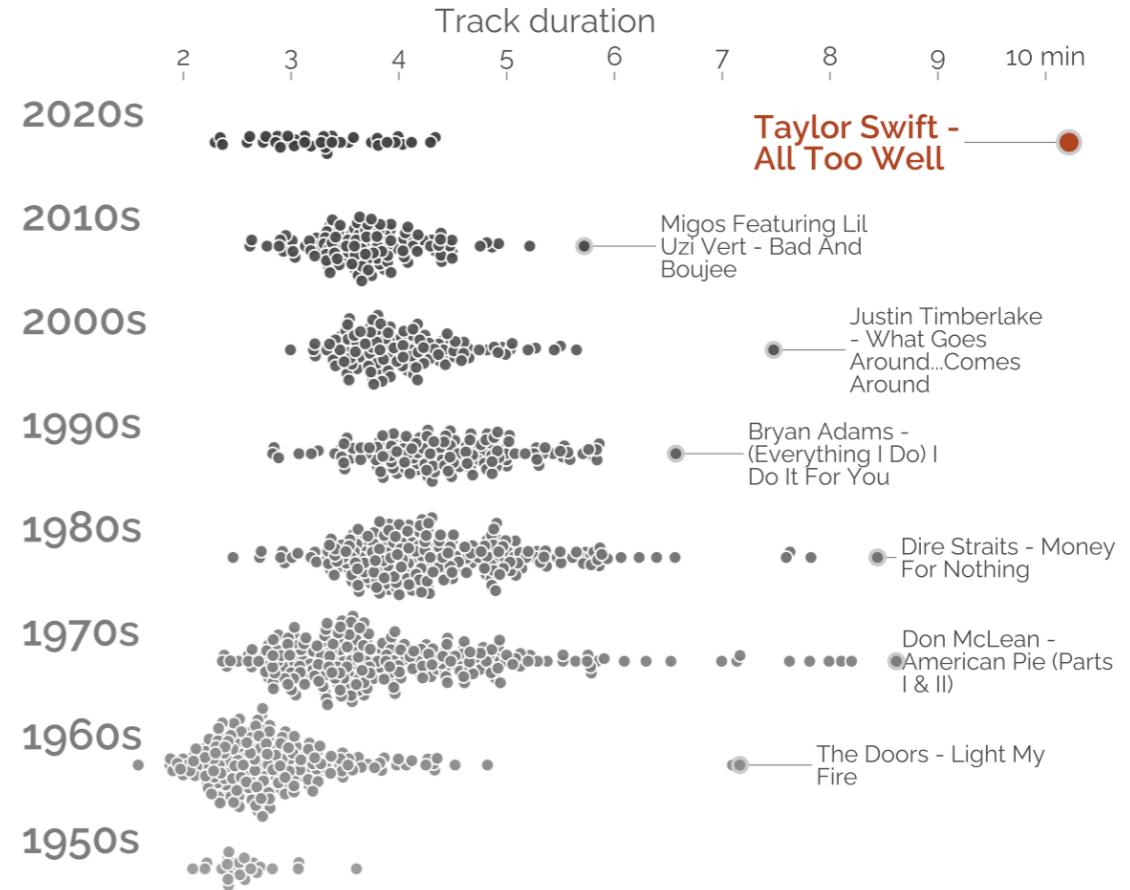
1 p <- p +
2   labs(
3     title = plot_titles$title,
4     subtitle = plot_titles$subtitle,
5     caption = plot_titles$caption,
6     y = "Track duration"
7   )

```

Not shown:
 The titles have been defined in
 the list object `plot_titles`

All Too Well is the Longest Song in U.S. Chart History

Taylor Swift's new version of "All Too Well" is the longest track to ever top the Billboard Hot 100 with a duration of 10 minutes and 13 seconds. The song replaces **Don McLean's song "American Pie"** (8 minutes and 36 seconds), which became no. 1 on 15 January 1972, after nearly 50 years. Each no. 1 song is a dot. The longest no. 1 song in each decade is highlighted.



Source: [Billboard Hot 100 \(Kaggle\)](#), [Spotify API](#) | Visualization: [Ansgar Wolsing](#)



Resources

- Talk at Data Visualization Meetup Munich March 2024:
github.com/bydata/talk-data-visualization-munich-20240305
- Full R code:
github.com/bydata/30DayChartChallenge/tree/main/2022/09
- {ggplot2} documentation: ggplot2.tidyverse.org
- Economist chart: bsky.app/profile/alexselbyb.bsky.social/post/3lqptoe76jk2e
- Rick Beato on YouTube: www.youtube.com/watch?v=aqHePc6Yu7w



If you want to stay in touch

