Cyclistic Bike-Share Case Study

Shrabana Nandi

November 13, 2025

Contents

1 BUSINESS TASK

Cyclistic, a Chicago-based bike-share program, aims to increase its number of annual members. This analysis explores how casual riders and annual members use Cyclistic bikes differently, with the goal of helping the marketing team design targeted strategies to convert casual riders into annual subscribers. The report summarizes key behavioral differences and identifies data-driven opportunities for strategic intervention.

---

2 DATA SOURCE

This analysis uses Divvy Bike Share trip data for **Q1 2019** and **Q1 2020**, provided by Motivate International Inc. under a public data license. The datasets include ride timestamps, bike types, user types, and station locations but exclude personally identifiable information for privacy reasons. Only Q1 data was used due to RStudio memory limits. Although Cyclistic is a fictional company, the dataset accurately represents real-world bike-share behavior.

---

3 DATA CLEANING & PREPARATION

The 2019 and 2020 datasets were standardized to align columns and formats, including renaming fields in the 2019 file to match the 2020 structure. Rows with missing values, negative ride durations, or internal test rides ("HQ QR") were removed. Timestamp fields were converted to POSIXct, and ride_length (minutes) was computed.

Additional engineered features include:

- Month, Day of Week, Hour
- Season (Winter, Spring, Summer, Fall)
- Weekday vs Weekend
- Ride Duration Category (short, medium, long, extended)
- Round-trip flag (start == end station)

The final cleaned dataset is reliable, consistent, and suitable for analysis.

---

4 ANALYSIS INSIGHTS

Ride Duration: Casual riders consistently take longer trips, while members ride briefly but more frequently — reflecting leisure vs. commuting behaviors.

Time Patterns: Members peak sharply during weekday commute hours (8–9 AM, 5–6 PM). Casual riders peak mid-day and weekends, supporting recreational usage.

Seasonality: Casual ridership rises dramatically in spring and summer, while members maintain stable usage year-round.

Spatial Patterns: Members start rides near business districts and transit hubs, whereas casual riders begin trips near parks, waterfronts, and tourist-heavy areas.

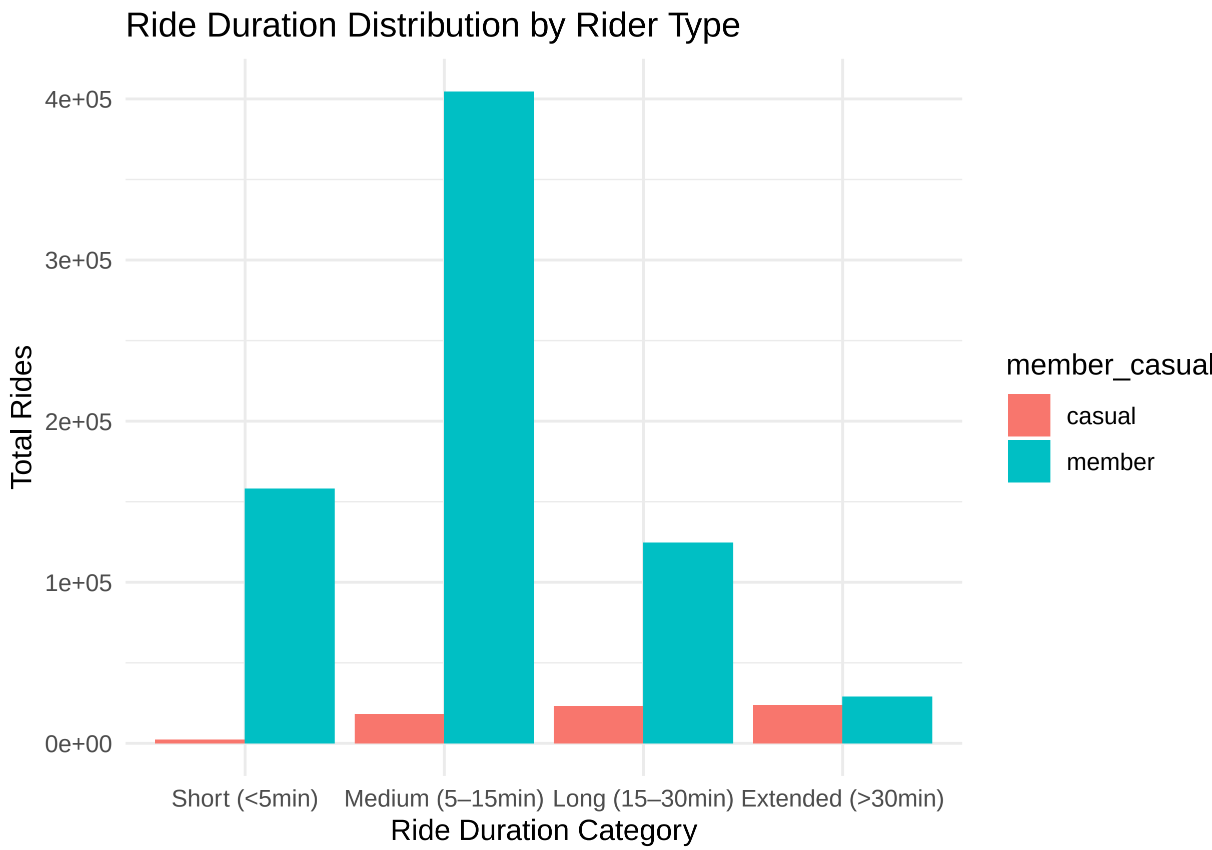
Trip Type: Casual riders take more round trips, while members show one-way commuting patterns.

These insights clarify the distinct motivations of both rider groups.

---

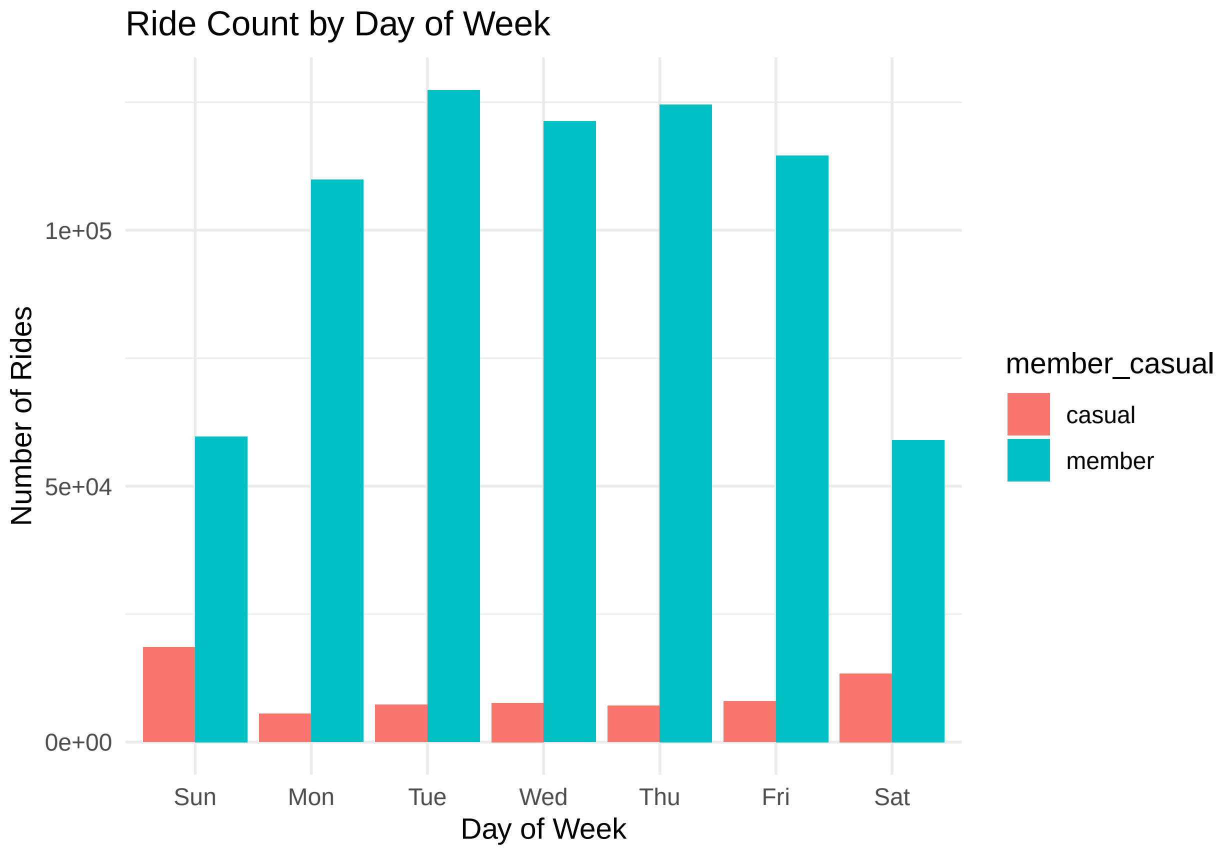
5 VISUALIZATIONS & KEY FINDINGS

5.1 Ride Duration Comparison



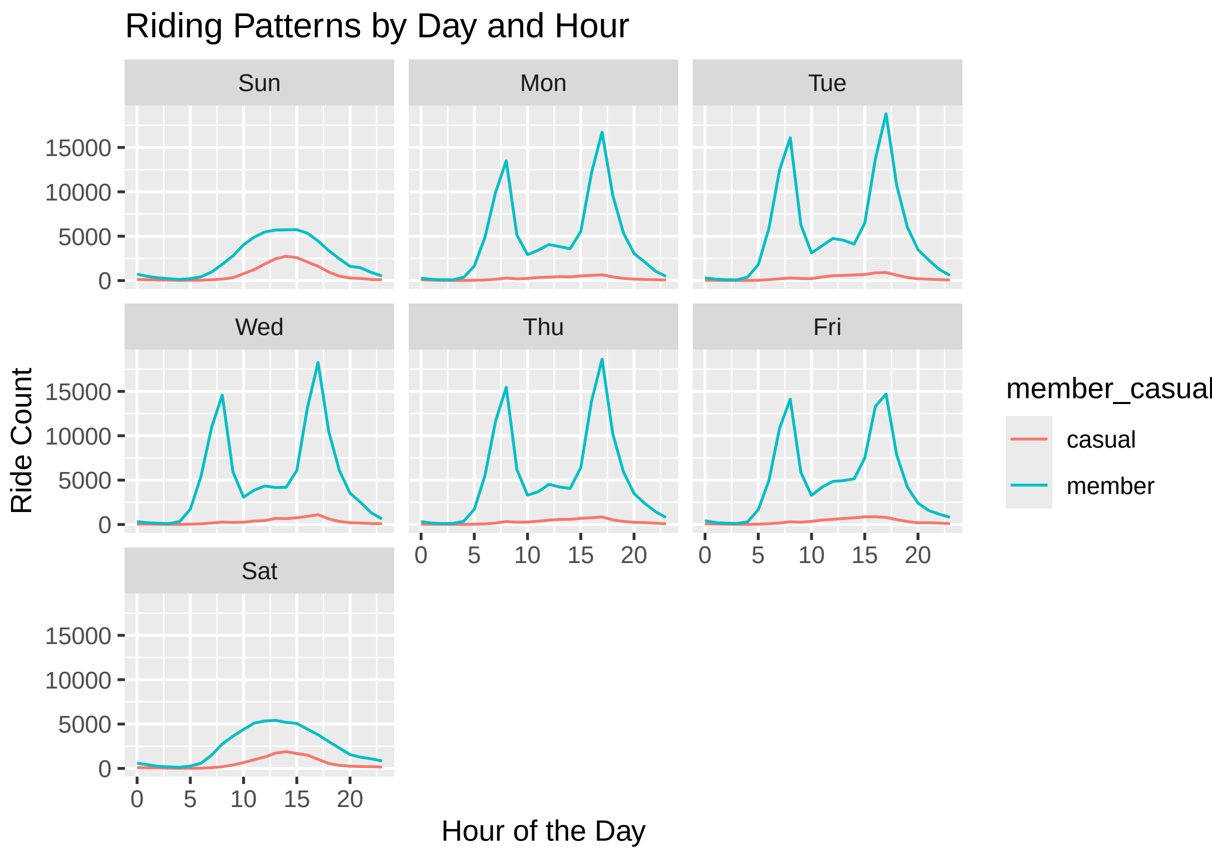
Casual riders have significantly longer ride durations than annual members.

5.2 Weekly Usage Patterns



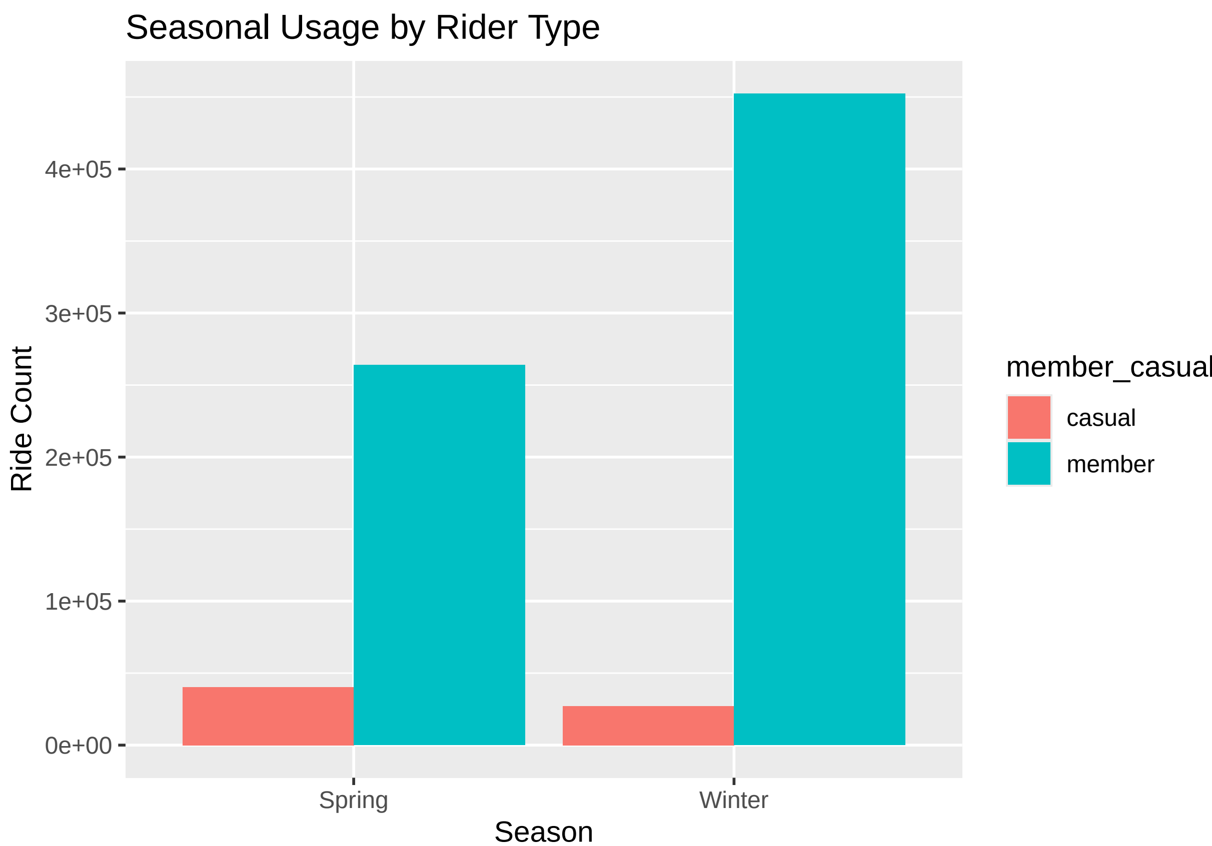
Members dominate weekdays; casual riders surge on weekends.

5.3 Hourly Trends by Day



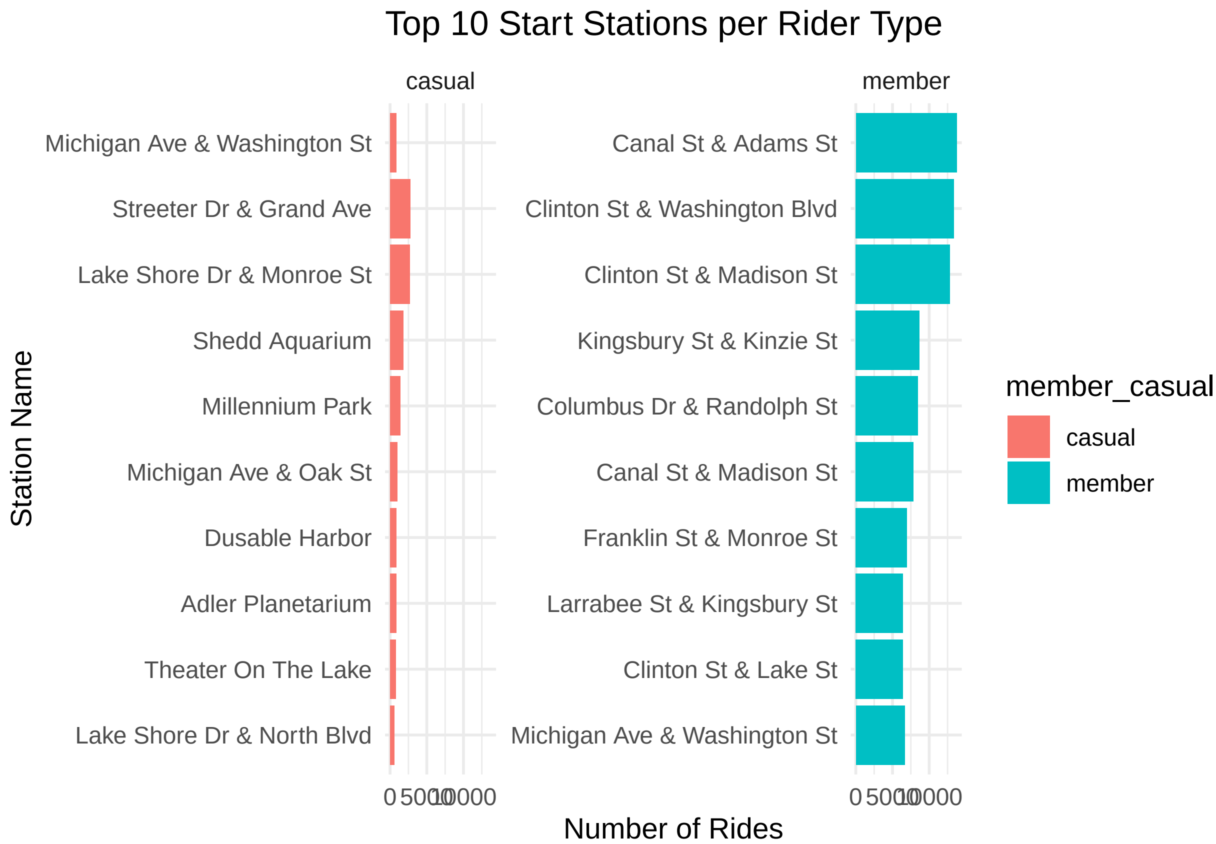
Members follow commute peaks; casual riders favor mid-day leisure periods.

5.4 Seasonal Trends



Casual ridership peaks in spring and summer.

5.5 Top Start Stations



Members cluster near workplaces; casual riders near scenic or tourist locations.

---

6 RECOMMENDATIONS

1. Seasonal Membership Campaign (Late Spring – Early Summer):

Cyclistic should launch focused membership promotions from late April to July, highlighting leisure-oriented perks such as ride discounts, scenic-station benefits, and attraction partnerships. Marketing, Partnerships, and Operations teams should collaborate to design and execute QR-based station ads.

2. Public Transport Advertising:

Membership ads should be placed inside buses, trains, and metro stations connected to tourist and weekend-heavy areas, using QR codes and simple messaging. Deployment should intensify from May to September. Marketing, Transit Advertising, and Brand Outreach teams will manage this initiative.

3. Extra Ride-Time Benefits for Members:

Introduce extended ride duration or monthly bonus minutes for annual members starting in June. This benefit aligns with casual riders' preference for long leisure rides and can be promoted by Product, Finance, and Marketing teams as part of membership onboarding.

---

7 CONCLUSION

The analysis shows clear and consistent differences between casual riders and annual members across time, location, and trip duration. Members use Cyclistic for structured, predictable commuting, while casual riders rely on it for leisure, exploration, and seasonal experiences. By focusing on seasonal promotions, well-placed advertising, and benefits aligned with casual rider behavior, Cyclistic can meaningfully increase annual membership conversion while enhancing user experience.

---

8 APPENDIX

8.1 A0. Libraries Used

```
#library("tidyverse")
#library("lubridate")
#library("janitor")
#library("skimr")
#library("readr")
#library("dplyr")
#library("ggplot2")
```

8.2 A1. Key Data Cleaning Steps

```
#Converting date-time columns
divvy_data <- divvy_data %>%
  mutate(
    started_at = as.POSIXct(started_at, format = "%Y-%m-%d %H:%M:%S"),
    ended_at = as.POSIXct(ended_at, format = "%Y-%m-%d %H:%M:%S")
  )

#Adding ride duration column
divvy_data <- divvy_data %>%
  mutate(ride_length = as.numeric(difftime(ended_at, started_at, units = "mins")))

#Standardize member_casual labels
divvy_data <- divvy_data %>%
  mutate(member_casual = recode(member_casual,
                                "Subscriber" = "member",
                                "Customer" = "casual"))

#Removing the long rides
divvy_data <- divvy_data %>%
  filter(ride_length > 0, ride_length < 1440)
```

8.3 A2. Summary Statistics

```
skim_without_charts(divvy_data)
```

Table 1: Data summary

Name	divvy_data
Number of rows	783799
Number of columns	15
Column type frequency:	
character	7
factor	2
numeric	4
POSIXct	2
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ride_id	0	1	7	16	0	783799	0
start_station_name	0	1	10	43	0	635	0
end_station_name	0	1	10	43	0	635	0
member_casual	0	1	6	6	0	2	0
season	0	1	6	6	0	2	0
week_type	0	1	7	7	0	2	0
ride_category	0	1	13	17	0	4	0

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
month	0	1	TRUE	3	Mar: 304307, Jan: 245745, Feb: 233747, Apr: 0
day_of_week	0	1	TRUE	7	Tue: 134628, Thu: 131653, Wed: 128925, Fri: 122542

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
start_station_id	0	1	202.03	155.85	2	77.00	174.00	289.00	673.00
end_station_id	0	1	202.01	156.24	2	77.00	173.00	289.00	673.00
ride_length	0	1	13.79	31.17	1	5.55	9.03	15.23	1435.92
hour	0	1	13.29	4.65	0	9.00	14.00	17.00	23.00

Variable type: POSIXct

skim_variable	n_missing	complete_rate	min	max	median	n_unique
started_at	0	1	2019-01-01 00:04:37	2020-03-31 23:51:34	2020-01-06 19:36:24	734957
ended_at	0	1	2019-01-01 00:11:07	2020-04-01 07:38:49	2020-01-06 19:48:48	730543

```r
summary(divvy_data$ride_length)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    1.000    5.550    9.033   13.790   15.233 1435.917
```

8.4 A3. Engineered Feature Definitions

- **ride_length:** difference between ended_at and started_at in minutes
- **season:** Winter (Dec–Feb), Spring (Mar–May), Summer (Jun–Aug), Fall (Sep–Nov)
- **week_type:** weekday vs weekend classification
- **ride_category:** ride duration grouped into Short, Medium, Long, Extended
- **is_round_trip:** TRUE when start and end stations are identical

---

8.5 A4. Environment Information

```r
sessionInfo()
```

```
## R version 4.5.1 (2025-06-13 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26200)
##
## Matrix products: default
##   LAPACK version 3.12.1
##
## locale:
## [1] LC_COLLATE=English_India.utf8  LC_CTYPE=English_India.utf8
## [3] LC_MONETARY=English_India.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_India.utf8
##
## time zone: Asia/Calcutta
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] rmarkdown_2.30  skimr_2.2.1     janitor_2.2.1   lubridate_1.9.4
##  [5] forcats_1.0.1   stringr_1.5.2   dplyr_1.1.4     purrr_1.1.0
##  [9] readr_2.1.5     tidyr_1.3.1     tibble_3.3.0    ggplot2_4.0.0
## [13] tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.6         generics_0.1.4     stringi_1.8.7      hms_1.1.4
##  [5] digest_0.6.37      magrittr_2.0.4     evaluate_1.0.5     grid_4.5.1
##  [9] timechange_0.3.0   RColorBrewer_1.1-3 fastmap_1.2.0      jsonlite_2.0.0
## [13] scales_1.4.0       cli_3.6.5          rlang_1.1.6        crayon_1.5.3
## [17] bit64_4.6.0-1      base64enc_0.1-3    withr_3.0.2        repr_1.1.7
## [21] yaml_2.3.10        tools_4.5.1        parallel_4.5.1     tzdb_0.5.0
## [25] vctrs_0.6.5        R6_2.6.1           lifecycle_1.0.4    snakecase_0.11.1
## [29] bit_4.6.0          vroom_1.6.6        pkgconfig_2.0.3    pillar_1.11.1
## [33] gtable_0.3.6       glue_1.8.0         haven_2.5.5        xfun_0.53
```

```
## [37] tidyselect_1.2.1   rstudioapi_0.17.1  knitr_1.50         farver_2.1.2
## [41] htmltools_0.5.8.1  labeling_0.4.3     compiler_4.5.1     S7_0.2.0
```

‘‘‘