

```

library(tidyverse) library(lubridate) library(janitor) library(skimr)

June_2021 <- read_csv("202106-divvy-tripdata.csv") July_2021 <- read_csv("202107-divvy-tripdata.csv")
August_2021 <- read_csv("202108-divvy-tripdata.csv") September_2021 <- read_csv("202109-divvy-tripdata.csv")
October_2021 <- read_csv("202110-divvy-tripdata.csv") November_2021 <- read_csv("202111-divvy-tripdata.csv")
December_2021 <- read_csv("202112-divvy-tripdata.csv") January_2022 <- read_csv("202201-divvy-tripdata.csv")
February_2022 <- read_csv("202202-divvy-tripdata.csv") March_2022 <- read_csv("202203-divvy-tripdata.csv")
April_2022 <- read_csv("202204-divvy-tripdata.csv") May_2022 <- read_csv("202205-divvy-tripdata.csv")

colnames(June_2021) colnames(July_2021) colnames(August_2021) colnames(September_2021) colnames(October_2021)
colnames(November_2021) colnames(December_2021) colnames(January_2022) colnames(February_2022)
colnames(March_2022) colnames(April_2022) colnames(May_2022)

compare_df_cols( June_2021, July_2021, August_2021, September_2021, October_2021, November_2021,
December_2021, January_2022, February_2022, March_2022, April_2022, May_2022 )

combined_trips <- bind_rows( June_2021, July_2021, August_2021, September_2021, October_2021,
November_2021, December_2021, January_2022, February_2022, March_2022, April_2022, May_2022 )

str(combined_trips) dim(combined_trips) head(combined_trips)

combined_tripsdate <- as.Date(combined_tripsstarted_at) combined_tripsmonth <- format(as.Date(combined_tripsdate),
"%B") combined_tripsday <- format(as.Date(combined_tripsdate), "%d") combined_tripsyear <-
format(as.Date(combined_tripsdate), "%Y") combined_tripsday_of_the_week <- weekdays(combined_tripsdate)

head(combined_trips)

combined_tripsride_length <- difftime(combined_tripsended_at, combined_tripsstarted_at, units =
"min") combined_tripsride_length <- round(combined_tripsride_length, 2) combined_tripsride_length <-
as.numeric(as.character(combined_trips$ride_length))

str(combined_trips)

combined_trips_clean <- drop_na(combined_trips)

combined_trips_clean <- filter(combined_trips_clean, ride_length>0)

lat_lng <- select(combined_trips_clean, start_station_name, end_station_name, start_lat, start_lng, end_lat, end_lng, member_id)
combined_trips_clean <- combined_trips_clean %>% select(-c(start_lat, start_lng, end_lat, end_lng))

head(combined_trips_clean)

combined_trips_agg_route <- unite(combined_trips_clean, "ride_route", start_station_name,
end_station_name, sep=" to ")

head(combined_trips_agg_route)

combined_trips_agg_routemonth <- ordered(combined_trips_agg_route$month, levels=c( "June", "July",
"August", "September", "October", "November", "December", "January", "February", "March", "April",
"May" ))

combined_trips_agg_routeday_of_the_week <- ordered(combined_trips_agg_route$day_of_the_week, levels=c(
"Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday" ))

combined_trips_agg_route %>% group_by(member_id) %>% summarise(number_of_rides=n(), average_ride_length=mean(ride_length))

combined_trips_agg_route %>% count(rideable_type)

combined_trips_agg_route %>% group_by(member_id) %>% count(rideable_type)

summary(combined_trips_agg_route$ride_length)

```

```

combined_trips_agg_route %>% group_by(member_casual) %>% summarise( number_of_rides =
n(), min_ride_length = min(ride_length), max_ride_length = max(ride_length), avg_ride_length =
mean(ride_length), median_ride_length = median(ride_length))

combined_trips_agg_route %>% group_by(member_casual, month) %>% summarise(number_of_rides=n(),
average_ride_length=mean(ride_length))%>% arrange (month)

combined_trips_agg_route %>% group_by(member_casual, day_of_the_week) %>% summarise(number_of_rides=n(),
average_ride_length=mean(ride_length))%>% arrange (day_of_the_week)

combined_trips_clean %>% group_by(start_station_name, member_casual) %>% summarise(number_of_trips=n())
%>% arrange(desc(number_of_trips)) %>% filter(member_casual=="casual") %>% select(start_station_name,
number_of_trips)

combined_trips_clean %>% group_by(end_station_name, member_casual) %>% summarise(number_of_trips=n())
%>% arrange(desc(number_of_trips)) %>% filter(member_casual=="casual") %>% select(end_station_name,
number_of_trips)

top_routes <- combined_trips_agg_route %>% group_by(ride_route) %>% summarise(number_of_rides=n())
%>% arrange(desc(number_of_rides))

head(top_routes)

top_routes_member_casual <- combined_trips_agg_route %>% group_by(ride_route, member_casual)
%>% summarise(number_of_rides=n()) %>% arrange(desc(number_of_rides))

head(top_routes_member_casual)

combined_trips_agg_route %>% group_by(member_casual) %>% summarise(average_ride_length =
mean(ride_length)) %>% ggplot(aes(x=member_casual, y=average_ride_length, fill=member_casual))+
geom_col() + labs(title="Average Ride length by rider type", x="Rider Type", y="Average Ride Length
(minutes)")

combined_trips_agg_route %>% group_by(member_casual, month) %>% summarise(number_of_rides=n(),
average_ride_length=mean(ride_length)) %>% ggplot(aes(x=month, y=number_of_rides, fill=member_casual))+
geom_col(position="dodge2") + labs(title="Number of rides per month by rider type", x="Month",
y="Number of rides") + theme(axis.text.x=element_text(angle=60, hjust=1))

combined_trips_agg_route %>% group_by(member_casual, month) %>% summarise(number_of_rides=n(),
average_ride_length=mean(ride_length)) %>% ggplot(aes(x=month, y=average_ride_length, fill=member_casual))+
geom_col(position="dodge2") + labs(title="Average ride length per month by rider type", x="Month",
y="Average ride length (minutes)") + theme(axis.text.x=element_text(angle=60, hjust=1))

combined_trips_agg_route %>% group_by(member_casual, day_of_the_week) %>% summarise(number_of_rides=n(),
average_ride_length=mean(ride_length)) %>% ggplot(aes(x=day_of_the_week, y=number_of_rides,
fill=member_casual)) + geom_col(position="dodge2") + labs(title="Number of rides per day of the week
by rider type", x="Day of the week", y="Number of rides") + theme(axis.text.x=element_text(angle=60,
hjust=1))

combined_trips_agg_route %>% group_by(member_casual, day_of_the_week) %>% summarise(number_of_rides=n(),
average_ride_length=mean(ride_length)) %>% ggplot(aes(x=day_of_the_week, y=average_ride_length,
fill=member_casual)) + geom_col(position="dodge2") + labs(title="Number of rides per day of the week by
rider type", x="Day of the week", y="Average ride length (minutes)") + theme(axis.text.x=element_text(angle=60,
hjust=1))

combined_trips_agg_route %>% group_by(rideable_type, member_casual) %>% summarise(number_of_rides=n())
%>% ggplot(aes(x=rideable_type, y=number_of_rides, fill=member_casual)) + geom_col(position="dodge2") +
labs(title="Number of rides per rideable type by rider type", x="Rideable type", y="Number of rides") +
theme(axis.text.x=element_text(angle=60, hjust=1))

combined_trips_agg_route %>% group_by(rideable_type, member_casual) %>% summarise(average_ride_length=mean(
%>% ggplot(aes(x=rideable_type, y=average_ride_length, fill=member_casual)) + geom_col(position="dodge2") +

```

```
labs(title="Average ride length per rideable type by rider type", x="Rideable type", y="Average ride length (minutes)") + theme(axis.text.x=element_text(angle=60, hjust=1))
```

```
combined_trips_agg_route %>% group_by(member_casual, month, rideable_type) %>% summarise(number_of_rides=n(), average_ride_length=mean(ride_length)) %>% ggplot(aes(x=month, y=number_of_rides, fill=rideable_type)) + geom_col(position="dodge2") + facet_wrap(~member_casual) + labs(title="Number of rides per month by rideable type for each rider type", x="Month", y="Number of rides") + theme(axis.text.x=element_text(angle=60, hjust=1))
```

```
combined_trips_agg_route %>% group_by(member_casual, day_of_the_week, rideable_type) %>% summarise(number_of_rides=n(), average_ride_length=mean(ride_length)) %>% ggplot(aes(x=day_of_the_week, y=number_of_rides, fill=rideable_type)) + geom_col(position="dodge2") + facet_wrap(~member_casual) + labs(title="Number of rides per day of the week by rideable type for each rider type", x="Day of the week", y="Number of rides") + theme(axis.text.x=element_text(angle=60, hjust=1))
```

```
combined_trips_clean %>% group_by(start_station_name, member_casual) %>% summarise(number_of_trips=n()) %>% arrange(desc(number_of_trips)) %>% filter(member_casual=="casual", number_of_trips>=15460) %>% select(start_station_name, number_of_trips) %>% ggplot(aes(x=start_station_name, y=number_of_trips)) + geom_col(fill="red") + coord_flip() + labs(title="Top 10 most popular start stations for casual riders", x="Start station name", y="Number of trips")
```

```
combined_trips_clean %>% group_by(end_station_name, member_casual) %>% summarise(number_of_trips=n()) %>% arrange(desc(number_of_trips)) %>% filter(member_casual=="casual", number_of_trips>=15460) %>% select(end_station_name, number_of_trips) %>% ggplot(aes(x=end_station_name, y=number_of_trips)) + geom_col(fill="green") + coord_flip() + labs(title="Top 10 most popular end stations for casual riders", x="End station name", y="Number of trips")
```

```
write.csv(top_routes, "C:\\Users\\User\\Documents\\Cyclistic Bike Sharing Data Analysis\\top_routes.csv", row.names=FALSE)
```

```
write.csv(combined_trips_agg_route, "C:\\Users\\User\\Documents\\Cyclistic Bike Sharing Data Analysis\\combined_trips_agg_route.csv", row.names=FALSE)
```

```
write.csv(top_routes_member_casual, "C:\\Users\\User\\Documents\\Cyclistic Bike Sharing Data Analysis\\top_routes_member_casual.csv", row.names=FALSE)
```