

Cyclistic Bike Share: Analytics

David Jackson

2021-08-12 17:40:25

Data Wrangling Change Log

August 6, 2021:

- Total Rows: 3,489,749
- I raised the following three concerns (about the data) with Lily Moreno(The director of marketing):
 - (1) 10,552 rows with negative trip durations,
 - (2) 122,175 rows with missing Starting Station names (and ID),
 - (3) As many as 1,291 stations with two different start_station_id.
- I was advised by Lily Moreno that we could ignore rows with trip duration ≤ 0 . We could also ignore rows with missing start_station_name, and also ignore stations_id.
- And should focus on doing using the start_station_name to perform aggregate functions on date, start_station_name, member_casual and rideable_type.
- Cleaning reduced rows to 3,343,689

Plot of Rides By Date

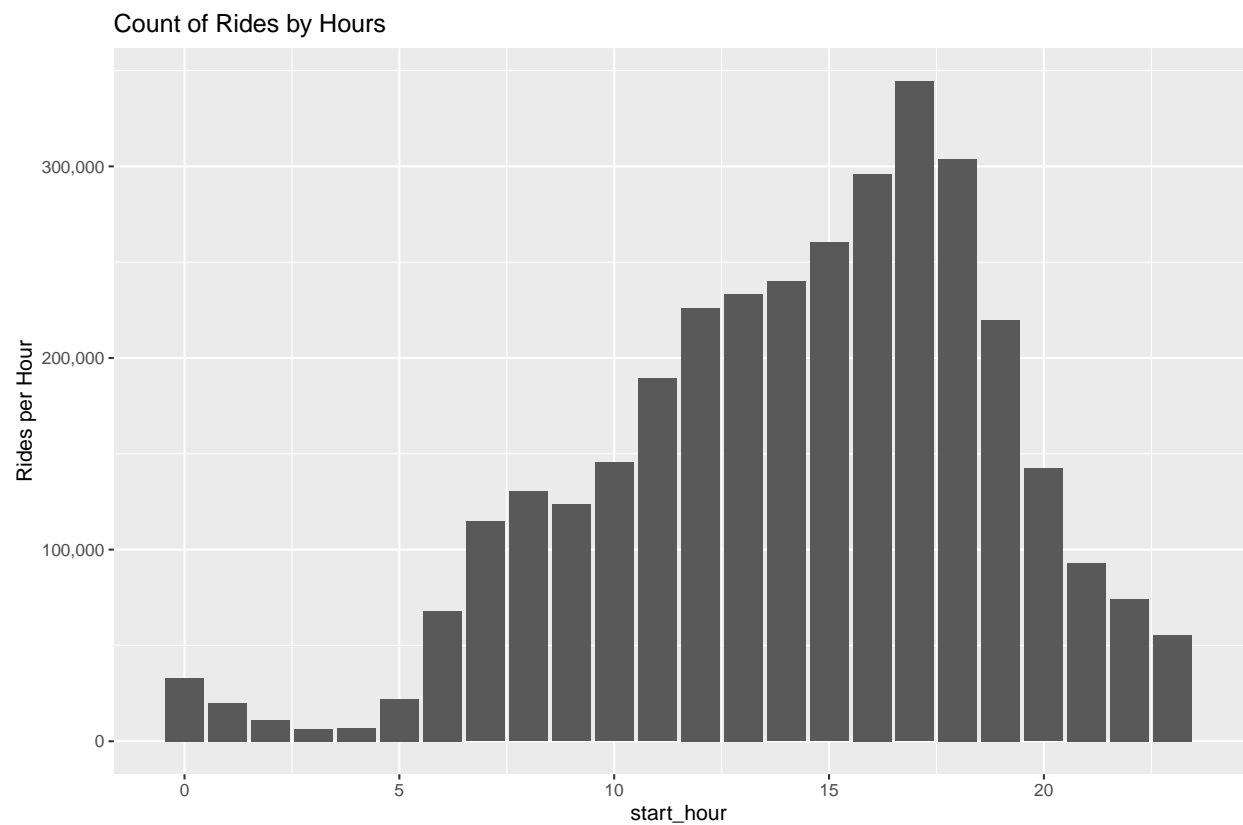
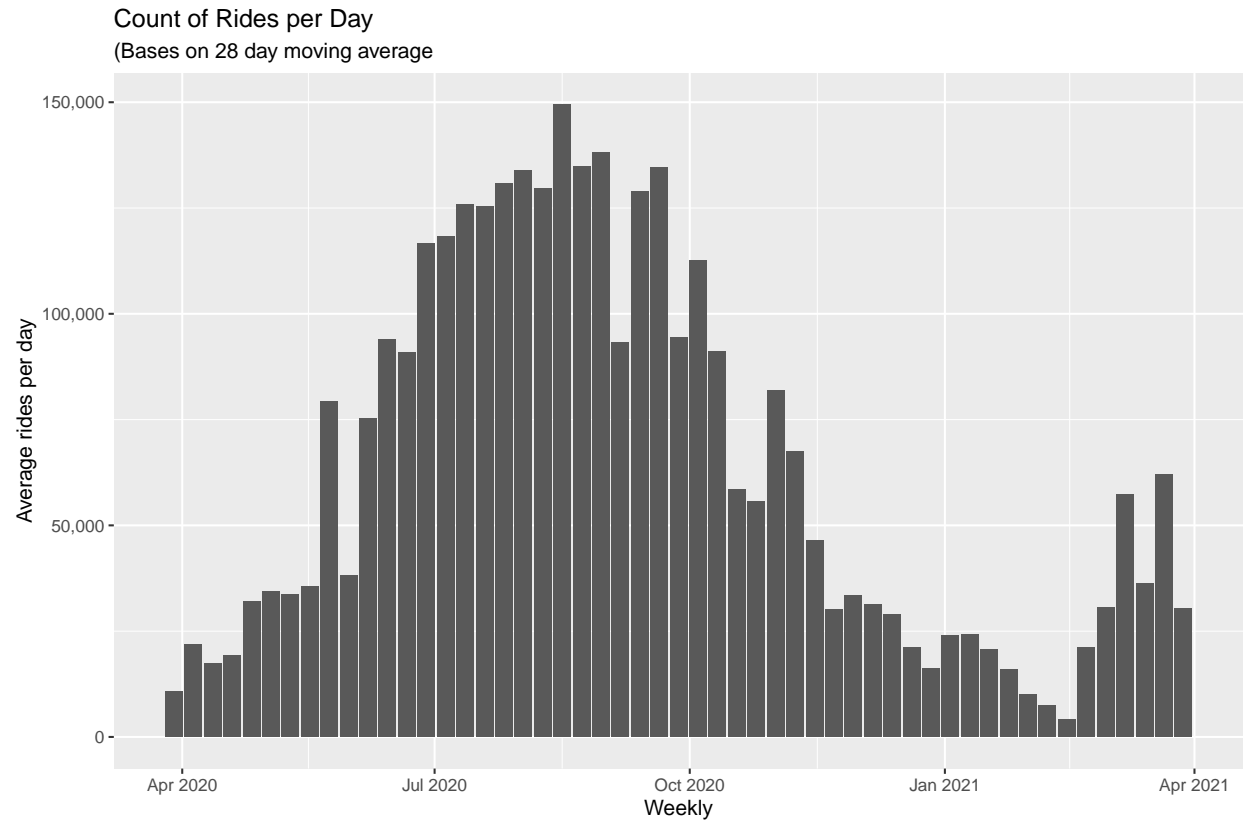
Summary Stats: Counts

- Summary of Hourly Counts

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	2.0	394.5	1475.0	2649.3	3779.5	15459.0

- Count of rides by Hour

##	bikesrides2\$start_hour										
##	0	1	2	3	4	5	6	7	8	9	10
##	32789	19708	10701	6016	6589	21916	67525	114563	130382	123473	145630
##	11	12	13	14	15	16	17	18	19	20	21
##	189354	225849	233341	239761	260297	296164	344403	303925	219467	142350	92908
##	22	23									
##	74043	55530									

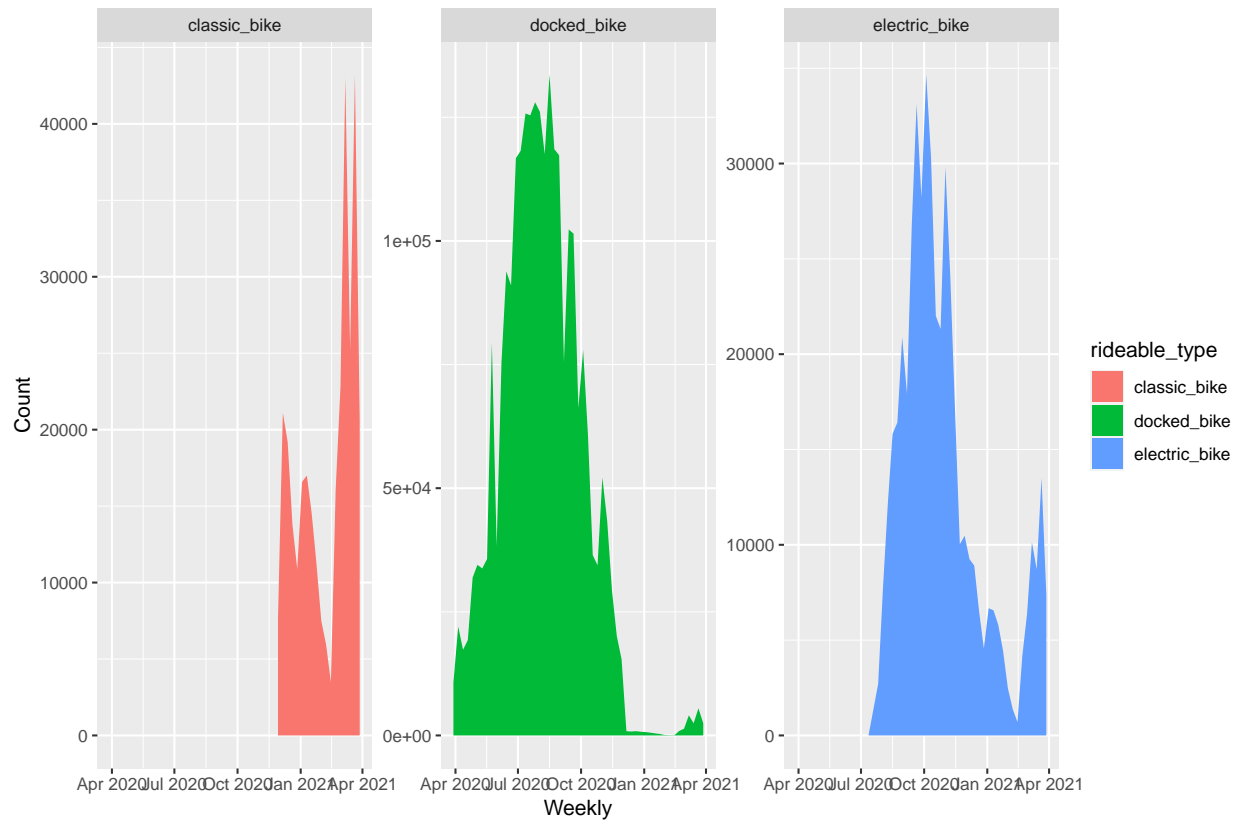


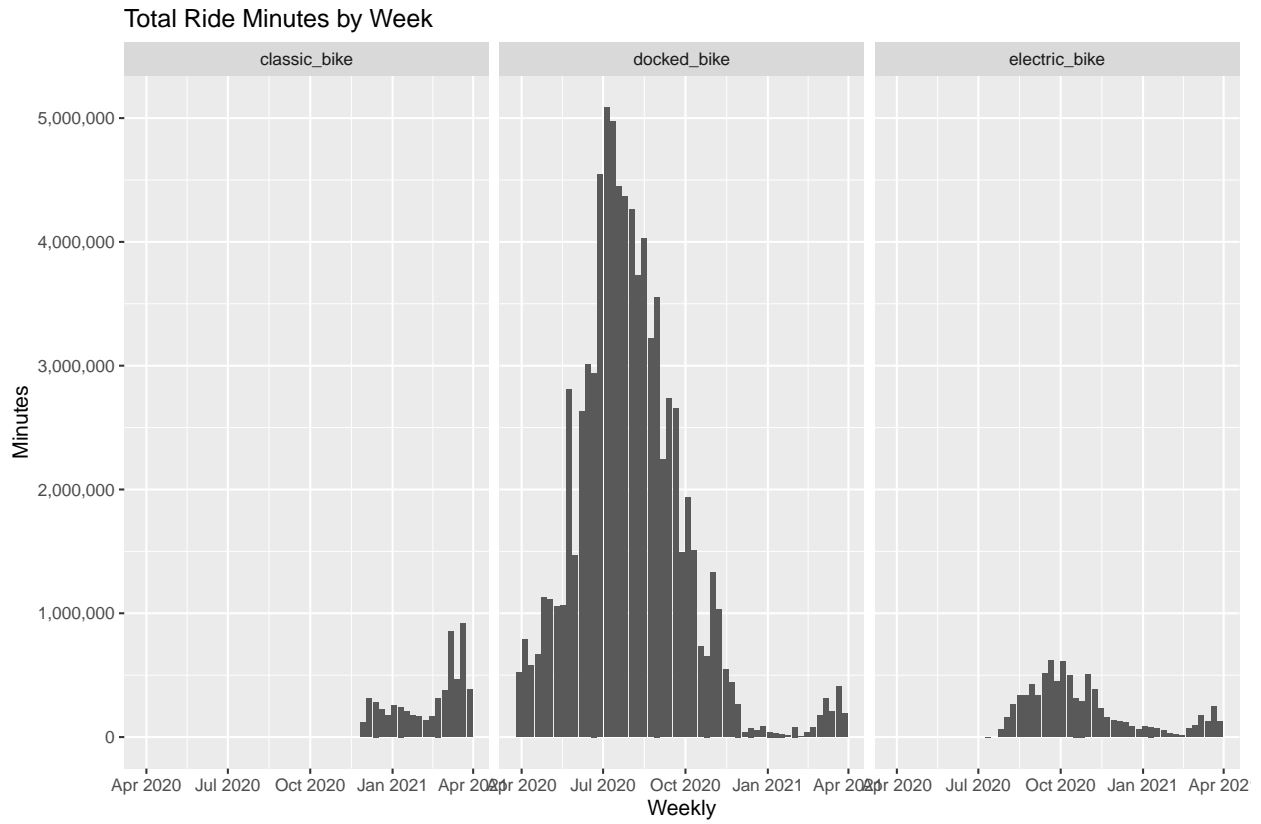
Count of Rides by Bike Type

Summary of Bike Types

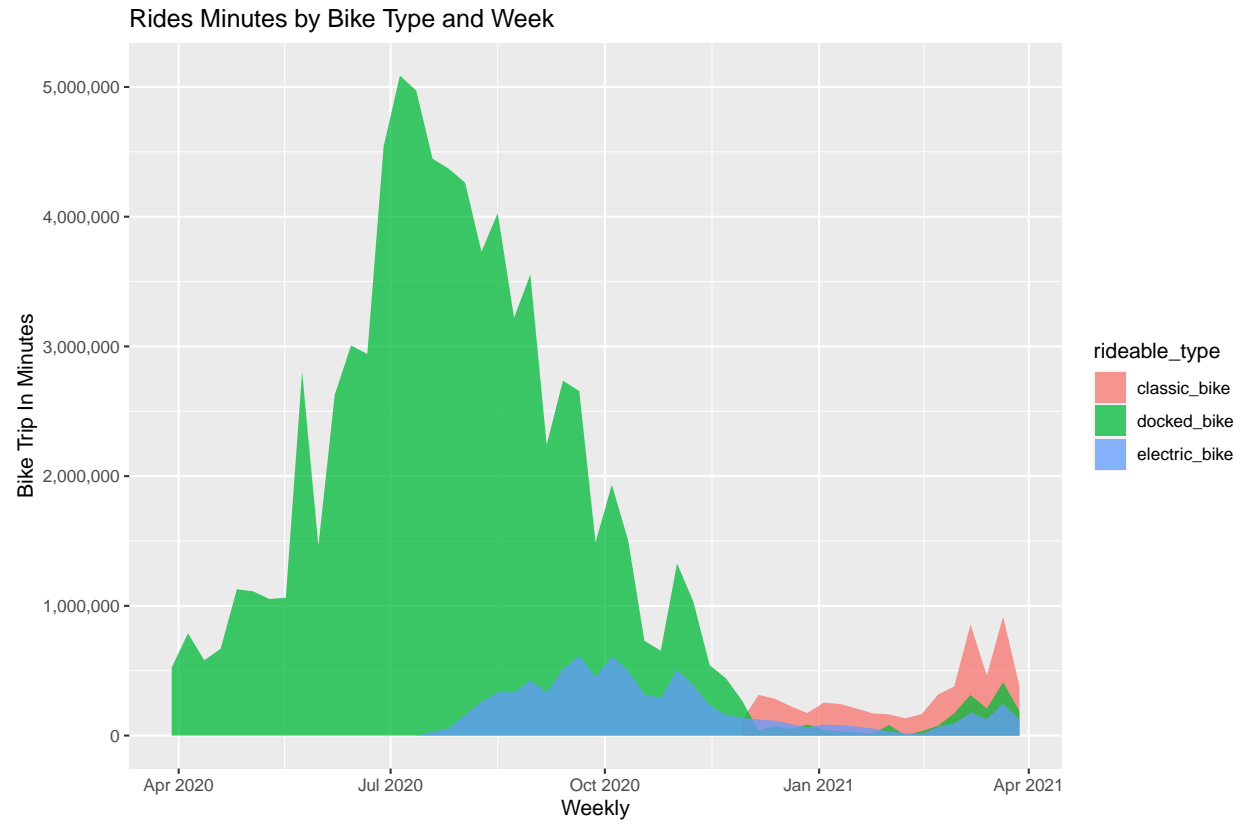
- Count by Bike Type(Total by Week)

```
##
## classic_bike   docked_bike electric_bike
##              18          53          37
```





```
## Warning: Width not defined. Set with 'position_dodge(width = ?)'
```



Begin Analysis Of Minutes of Ride Time

