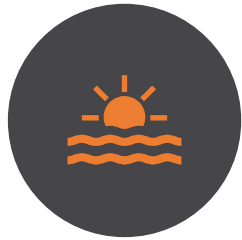




Walt Disney World Ride Data

By Hailey Skoglund, Gus Lipkin, Alex Pierstorff

Research Questions



Does hot weather increase wait times for rides with a "splash aspect"?



Does rainy or hot weather cause an increase in wait times for "indoor rides"?



Determine the busiest day of week and busiest month of the year (lubridate data) for tourists to visit WDW.



Create a structured travel plan for tourists to minimize the amount of time spent waiting in line.

```

####{r}
dwarfs_train <- read_csv("../data/7_dwarfs_train.csv")
alien_saucers <- read_csv("../data/alien_saucers.csv")
dinosaur <- read_csv("../data/dinosaur.csv")
expedition_everest <- read_csv("../data/expedition_everest.csv")
flight_of_passage <- read_csv("../data/flight_of_passage.csv")
kilimanjaro_safaris <- read_csv("../data/kilimanjaro_safaris.csv")
navi_river <- read_csv("../data/navi_river.csv")
pirates_of_caribbean <- read_csv("../data/pirates_of_caribbean.csv")
rock_n_rollercoaster <- read_csv("../data/rock_n_rollercoaster.csv")
slinky_dog <- read_csv("../data/slinky_dog.csv")
soarin <- read_csv("../data/soarin.csv")
spaceship_earth <- read_csv("../data/spaceship_earth.csv")
splash_mountain <- read_csv("../data/splash_mountain.csv")
toy_story_mania <- read_csv("../data/toy_story_mania.csv")

```

```

####{r, message=FALSE, warning=FALSE}
library(tidyverse)
Disney_Data <- read_csv("../data/metadata.csv")
glimpse(Disney_Data)

```

```

####{r}
metadata <- read_csv("../data/metadata.csv")
head(metadata)

```

DATE	WDW_TICKET_SEASON	DAYOFWEEK	DAYOFYEAR	WEEKOFYEAR	MONTHOFYE...	YEAR	SEASON	HOLIDAYPX	HOLIDAYM
1/1/12	NA	1	0	1	1	2012	CHRISTMAS PEAK	0	5
1/2/12	NA	2	1	1	1	2012	CHRISTMAS	2	5
1/3/12	NA	3	2	1	1	2012	CHRISTMAS	3	0
1/4/12	NA	4	3	1	1	2012	CHRISTMAS	4	0
1/5/12	NA	5	4	1	1	2012	CHRISTMAS	5	0
1/6/12	NA	6	5	1	1	2012	CHRISTMAS	6	0

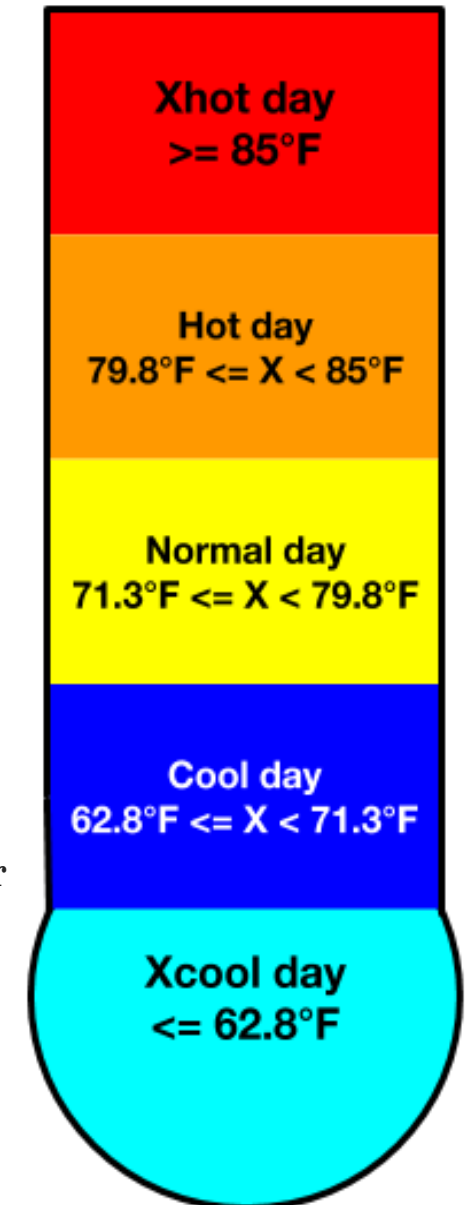
6 rows 11-10 of 190 columns

Data Involved

- Lots of Datasets!
- Metadata includes:
 - Dates
 - Ticket sales seasons
 - Holidays
 - School sessions nearby
- Rides_df includes:
 - Information on each of the 14 rides analyzed
 - Compilation of each individual ride's dataframe
- Individual Rides: 14 dataframes
 - Individual ride's names
 - Wait times
 - Reported , actual, and average wait times
 - date

Analysis Performed

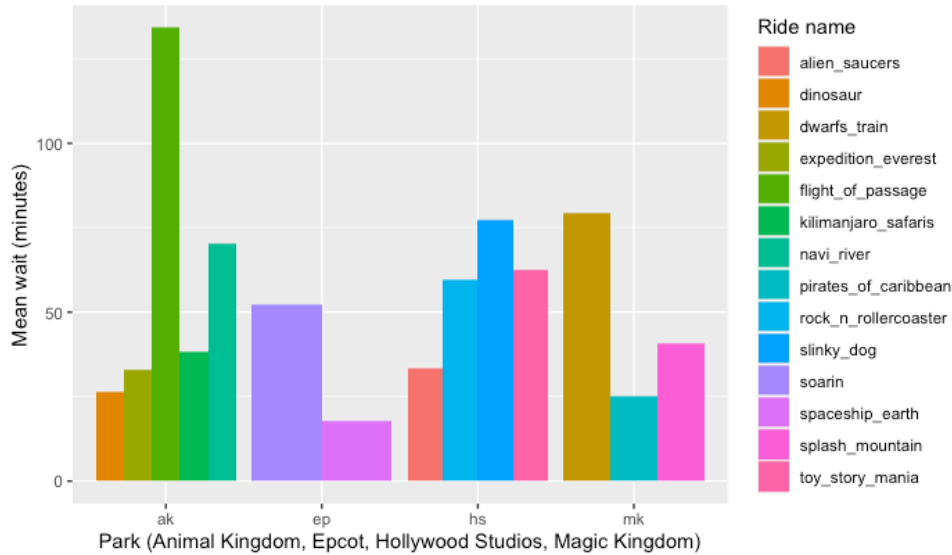
- Temperatures
 - Organized temps into categories based on mean temperatures
- Parade Times
 - Visualization showing wait times in proximity to parade start times.
- Rainfall
 - Explored relationship between average precipitation and wait times.
- Busiest Seasons
 - Used lubridate package to identify the busiest dates of the week, month, and year
- Holiday Proximity
 - Analyzed peaks in wait times in relation to proximity of seasonal holiday events.



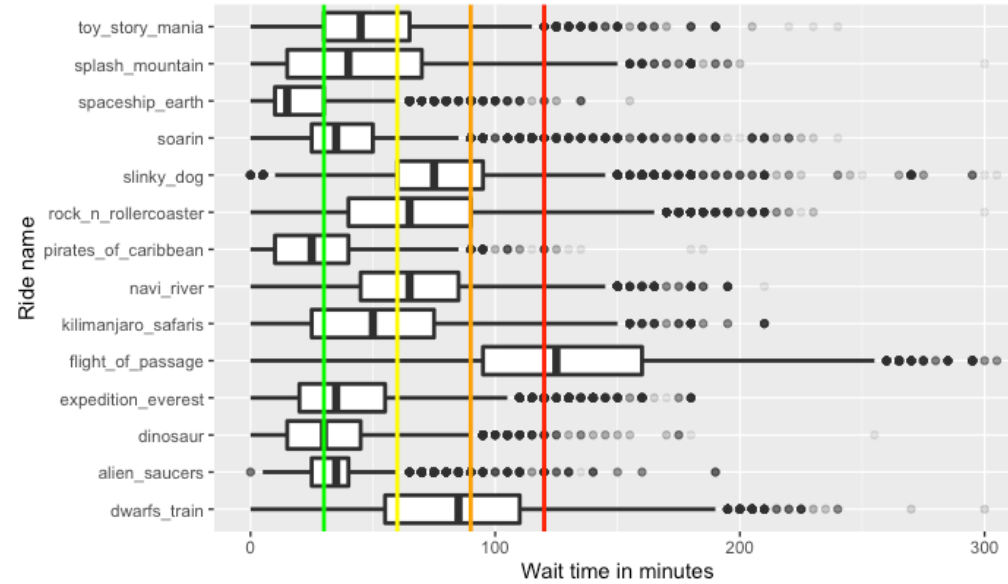
Visualizations

- Wait times are recorded as either park reported estimates or user reported actual times

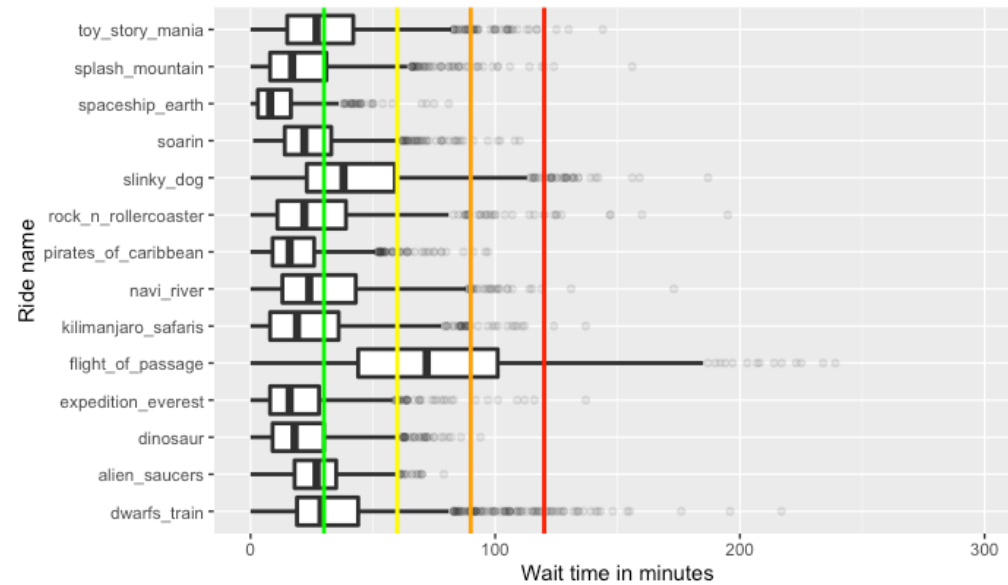
Column chart of mean estimated wait time by ride, grouped by park



Boxplot of mean estimated wait time by ride

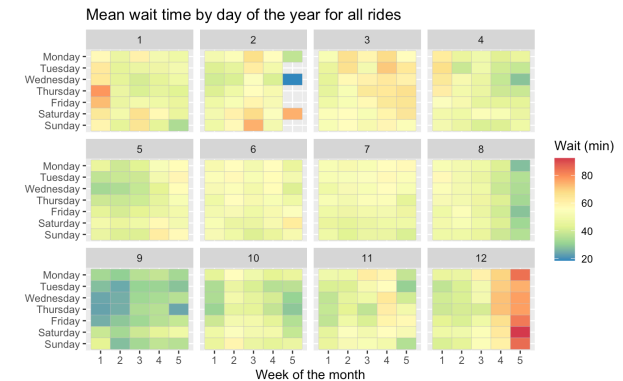
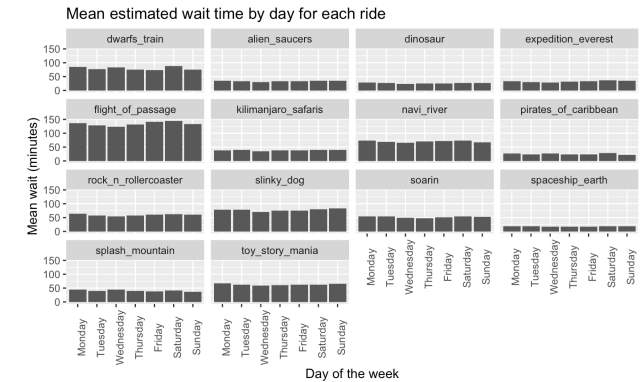
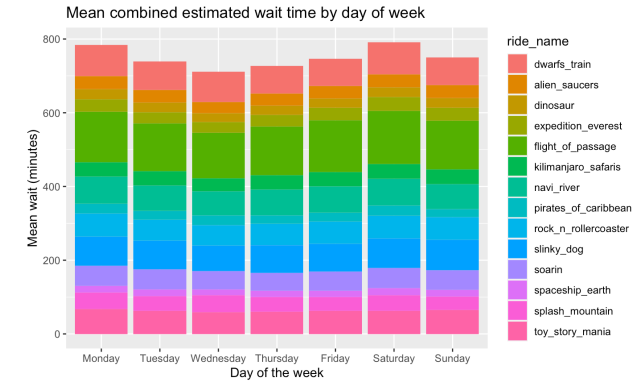


Boxplot of mean user wait time by ride



Visualizations

- Bar graphs
 - Shows the average wait times of each of the rides throughout the week.
 - Compares wait times of each ride for each day of the week in a side-by-side visualization.
- Tile graph
 - Shows which week of each month is the busiest by showing



Summary & Conclusion

- **Correlations:**
 - Temperature and types of rides:
 - Hot days and inside air-conditioned rides
 - Cold days and water rides
 - Rainy Weather
 - Increases wait times
 - Seasonal High's and Low's:
 - September return to school
 - December Holiday season
 - Parades
 - Slight increase in wait times due to crowds following parades
 - New Attractions
 - High wait times associated with opening days
- **Optimal Travel Plan:**
 - September
 - Wednesday
 - Epcot
 - Spaceship Earth first, then Soarin'

Next Steps

- **Linear Regression:**

- Allows us to directly compare how different factors affect wait times
- We can look at the charts we made and see which variables have the biggest affect and use those as wait time predictors

- **Example regression:**

- Are estimated wait times a good predictor of actual wait times?
- We filter where there is an estimated and user recorded wait time during the same minute and then get only unique estimated values
- Adjusted $r^2 = .4836$
- p-value = $3.012e^{-6}$
- **Conclusion:**
 - Not particularly based on this data

