Analysis of travel runs in Yellowstone

Natalia Brown

10/24/2017

Executive summary

Yellowstone National Park is visited every year by thousands of people from all over the country (and world). One the roads in the park connects the West Gate Entrance (referred to as the West Entrance) of the park to Old Faithful. The interest of this project is to identify locations in this road were major slowdows occur, and also estimate the travel times from the starts of the West entrance to Old Faithful. In order to accomplish this, travel time runs were done from these two locations in which the time and speed was recorded at every second. With these data, the speeds everywhere along the route were known and the travel times from one point of the road to the other could be calculated During the recording of the travel runs, the driver experienced animal traffic jams so it was also of interest to compare the travel times between an average peak day and this special day.

The travel runs were done on July 23rd -25th and September 3rd-5th. The wild day, the day of the animal traffic jam, was on July 25th. The peak days correspond to July 23rd, 24th, and September 3rd. The story to communicate is the location of major slowdowns for the peak day in the southbound direction, as well as showing the travel times of each run by time of day (AM or midday) in the southbound direction and comparing the average travel times between the average peak day and the wild day by time of day (AM or midday). As this was the purpose of the graphic, the final graphic shows the road in the analysis displaying the speeds at every point during the peak day runs, and other two graphs: one showing the travel times of each run by time of day, and another one showing the average travel time of the peak days and the travel time of the wild day by time of day. With these three graphics, the final graphics shows a complete story of what is happening on the southbound direction of the road connecting the West entrance of the park and Old Faithful.

Data background

Fehr & Peers is a transportation planning and engineering consulting firm. This firm is currently working on a project for the Yellowstone National Park, in which travel time runs were done during July and September of 2017. These travel runs consist of a person driving from one point to another and recording the time and speed at every second. This was done at various times of the day to analyze the patterns. Travel runs were done on both southbound and northbound directions. Each travel run was stored as a separate excel file containing a row for every second with its time, speed, longitude and latitude, and other variables that are not relevant to the project.

The travel time runs were done on July 23rd through the 25th (Sunday through Tuesday) and September 3rd through 5th (Sunday through Tuesday). While recording the travel time runs on July 25th, the driver experienced animal traffic jams, which meant that the travel times would be much higher than an average peak day, and there would be some spots with lower speeds (close to 0 miles per hour). This day was called the "wild day". With the data available, the purpose of the project was to summarize the travel times and speeds by direction and time of day, as well as comparing the travel times on an average peak day with the wild day travel times.

Data cleaning

The data of each travel run was stored as an excel file. The runs were previously classified by direction and time of day. The time of day of interest were AM and midday, which correspond to 8am-11am and 11am-2pm. Some runs ended at 11:30 so these runs were clasified as AM runs.

All the travel time runs had the same structure: the time, speed (mph), heading, latitude, longitude, HDOP, quality, # Sat and start time. The table below show a sample of an excel file containing the data for one run.

```
library(readxl)

table_sample <- head(read_excel("data/July/Grand Loop Rd 072317 RUN1 DS.xls")
,n=4)

knitr::kable(table_sample)
```

Time	Speed (MPH)	Heading	Latitude	Longitude	HDOP	Quality	# Sat	Start Time
1899-12-31 07:57:38	0.0	219.5	44.65877	-111.0984	0.8	2	9	07:51:21
1899-12-31 07:57:39	3.1	187.5	44.65876	-111.0984	0.8	2	9	07:51:21
1899-12-31 07:57:40	4.1	164.7	44.65875	-111.0984	0.8	2	9	07:51:21
1899-12-31 07:57:41	5.1	153.6	44.65873	-111.0984	0.8	2	9	07:51:21

The data was collected on July and September of 2017. To keep the data organized, a folder was created to store the July and September runs separately. The data had to be compiled by direction and time of day, therefore since the runs were previously clasiffied by those

parameters, the data only had to be loaded into R and combined. The runs corresponding to the peak days were loaded first followed by the wild day. Once the data was loaded into R, it had to be cleaned up since the travel times of all the runs had to be measured from the West Entrance of the park to Old Faithful (the data contained points outside of the area of interest). In order to do this, the data had to be filtered to delete all the points that were not part of the route. This involved selecting a reference point for each end of the route and filtering by longitude and latitude. Once that was done, it was also noticed that some points were not deleted and another condition was included in the filter. This last condition depended on the time at which the driver reached the end point of the route, so it varies by run. The filter applied to all the southbound runs was to keep all the runs within the two perviously defined points, and all the points recorded before the driver reached the end point of the route. Also, to better work with the data, only the columns relevant to the analysis were kept, and the variable names were changed. The variables kept were time, speed, latitude, and longitude. Other variables were added to the data frame: run, period (AM or MD), an direction (SB or NB).

```
#Load libraries
library(tidyverse)
library(sf)
library(ggmap)
library(lubridate)
library(forcats)

#Load SB AM runs for peak day

run1_SB_AM <- read_excel("data/July/Grand Loop Rd 072317 RUN1 DS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 1 SB AM") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
08:45:33"))

run2_SB_AM <- read_excel("data/July/Grand Loop Rd 072317 RUN3 DS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 2 SB AM") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
11:14:15"))

run3_SB_AM <- read_excel("data/July/Grand Loop Rd 072317 RUN6 PS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 3 SB AM") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
09:42:37"))

run4_SB_AM <- read_excel("data/July/Grand Loop Rd 072417 RUN2 DS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
```

```
  mutate(run = "Run 4 SB AM") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
11:19:00"))

run5_SB_AM <- read_excel("data/July/Grand Loop Rd 072417 RUN4 PS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 5 SB AM") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
10:43:04"))

#Combine all the SB AM runs into one data frame.

all_SB_AM <- bind_rows(run1_SB_AM, run2_SB_AM, run3_SB_AM, run4_SB_AM, run5_S
B_AM) %>%
  mutate(period = "AM", direction = "SB")
```

The same process was followed for the southbound midday runs. While loading all the runs corresponding to the southbound midday category, it was noticed that two runs were incomplete, and therefore were not included in the analysis. It was also found that one of the July 24th runs contained data with the wrong longitude and latitude, and therefore those data points were deleted.

```
#Load SB midday runs for peak day

run1_SB_MD <- read_excel("data/July/Grand Loop Rd 072317 RUN5 DS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 1 SB MD")  %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
12:52:15"))

run2_SB_MD <- read_excel("data/July/Grand Loop Rd 072317 RUN8 PS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 2 SB MD") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
11:42:18"))

#There are a few points that have the wrong latitude and longitude, so those
are filtered out
run3_SB_MD <- read_excel("data/July/Grand Loop Rd 072417 RUN6 PS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  filter(!lat == -1) %>%
  mutate(run = "Run 3 SB MD") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
12:30:17"))
```

```
#Combine all the SB midday runs into one data frame.

all_SB_MD <- bind_rows(run1_SB_MD, run2_SB_MD, run3_SB_MD)%>%
  mutate(period = "MD", direction = "SB")
```

The same process was followed to load all the northbound runs. However, these are not used for this particular visualization.

```
#Load NB AM runs for peak day

run1_NB_AM <- read_excel("data/July/Grand Loop Rd 072317 RUN2 DS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 1 NB AM")

run2_NB_AM <- read_excel("data/July/Grand Loop Rd 072317 RUN7 PS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 2 NB AM")

run3_NB_AM <- read_excel("data/July/Grand Loop Rd 072417 RUN1 DS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 3 NB AM")

run4_NB_AM <- read_excel("data/July/Grand Loop Rd 072417 RUN5 PS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 4 NB AM")

#There are a few points that have the wrong latitude and longitude, so those
are filtered out
run5_NB_AM <- read_excel("data/Sept/GLR 090317 RUN1.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  filter(!lat == -1) %>%
  mutate(run = "Run 5")

#Combine all the NB AM runs into one data frame.

all_NB_AM <- bind_rows(run1_NB_AM, run2_NB_AM, run3_NB_AM, run4_NB_AM)%>%
  mutate(period = "AM", direction = "NB")

#Load NB midday runs for peak day

run1_NB_MD <- read_excel("data/July/Grand Loop Rd 072317 RUN4 DS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 1 NB MD")
```

```
run2_NB_MD <- read_excel("data/July/Grand Loop Rd 072417 RUN7 PS.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run 2 NB MD")

#Combine all the NB midday runs into one data frame.

all_NB_MD <- bind_rows(run1_NB_MD, run2_NB_MD)%>%
  mutate(period = "MD", direction = "NB")

#Load wild day runs

WD_SB_AM <- read_excel("data/July/Grand Loop Rd 072517 RUN1.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run WD SB AM", period = "AM", direction = "SB") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
10:07:34"))

WD_SB_MD <- read_excel("data/July/Grand Loop Rd 072517 RUN3.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run WD SB MD", period = "MD", direction = "SB") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time < ymd_hms("1899-12-31
12:13:54"))

WD_NB_AM <- read_excel("data/July/Grand Loop Rd 072517 RUN2.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run WD NB AM", period = "AM", direction = "NB") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time > ymd_hms("1899-12-31
10:10:13"))

WD_NB_MD <- read_excel("data/July/Grand Loop Rd 072517 RUN4.xls") %>%
  select(time = Time, speed = `Speed (MPH)`, lat = Latitude, long = Longitude
) %>%
  mutate(run = "Run WD NB MD", period = "MD", direction = "NB") %>%
  filter(long > -111.0972444 & lat > 44.46177222 & time > ymd_hms("1899-12-31
12:16:49"))

#Combine all the wild day runs into one data frame.

wildday  <- bind_rows(WD_SB_AM, WD_SB_MD, WD_NB_AM, WD_NB_MD) %>%
  mutate(type_day = "Wild Day")
```

To better organize the data, all the peak day runs were combined into one master data frame. From that data frame, other data frames were created such as a data frame for the southbound and northbound directions (filtering by the "direction" variable in the data).

```
#Combine all the "all" files for the peakday to have one combined file to wor
k with

peakday <- bind_rows(all_SB_AM, all_SB_MD, all_NB_AM, all_NB_MD)%>%
  mutate(type_day = "Peak Day")

# Separate SB and NB for peak days

peak_SB_allday <- peakday %>%
  filter(direction == "SB")

peak_NB_allday <- peakday %>%
  filter(direction == "NB")

# Separate SB and NB for wild day

wd_SB_allday <- wildday %>%
  filter(direction == "SB")

wd_NB_allday <- wildday %>%
  filter(direction == "NB")
```

Once all the data was stored in a concise data frame by type of day and direction, the travel times were calculated. Since the data included the variable "run", the travel time was calculated by subtracting the first time stamp of each run to the last time stamp of the same run. This value was recorded in seconds and minutes for each run (peak and wild day).

```
#Calculate travel times for peak day
peakday_tt <- peakday %>%
  group_by(run, period, direction) %>%
  summarize(travel_time = difftime(last(time), first(time), units = "secs"))
%>%
  mutate(seconds= as.numeric(travel_time),
         minutes = as.numeric(travel_time)/60) %>%
  ungroup()

#Calculate travel times for wild day
wildday_tt <- wildday %>%
  group_by(run, period, direction) %>%
  summarize(travel_time = difftime(last(time), first(time), units = "secs"))
%>%
  mutate(seconds= as.numeric(travel_time),
         minutes = as.numeric(travel_time)/60) %>%
  ungroup()
```

Once the travel times were calculated for each run, these had to be summarized by direction (southbound and northbound) as well as by period (AM and midday).

```r
#Average travel times for peak day by direction and time of day

peakday_summary_tt <- peakday_tt %>%
  group_by(period, direction) %>%
  summarize(ave_tt = mean(minutes),
            sd_tt = sd(minutes))

#Separate SB travel times for plotting purposes

peakday_sb_tt_runs <- peakday_tt %>%
  filter(direction == "SB") %>%
  arrange(period) %>%
  mutate(travel_time = as.numeric(travel_time))

peakday_sb_tt <- peakday_summary_tt %>%
  filter(direction == "SB")

wildday_sb_tt <- wildday_tt %>%
  filter(direction == "SB")

# Separate the SB AM runs from the SB midday runs and create a "blank" row of
data to plot the graph and keep them separated

top <- peakday_sb_tt_runs %>%
  filter(period == "AM")

bottom <- peakday_sb_tt_runs %>%
  filter(period == "MD")

middle <- data_frame(run = c("1"),
                     period = c("1"),
                     direction = c("1"),
                     travel_time = 0,
                     seconds = 0,
                     minutes = 0)

#Combine the SB travel time data frames to have a nice graph that separates t
he AM from the midday runs

peakday_sb_tt_v2 <- bind_rows(top, middle, bottom) %>%
  mutate(run = fct_inorder(run,ordered = TRUE))
```

Individual figures

Figure 1: Peak day Speed Distribution

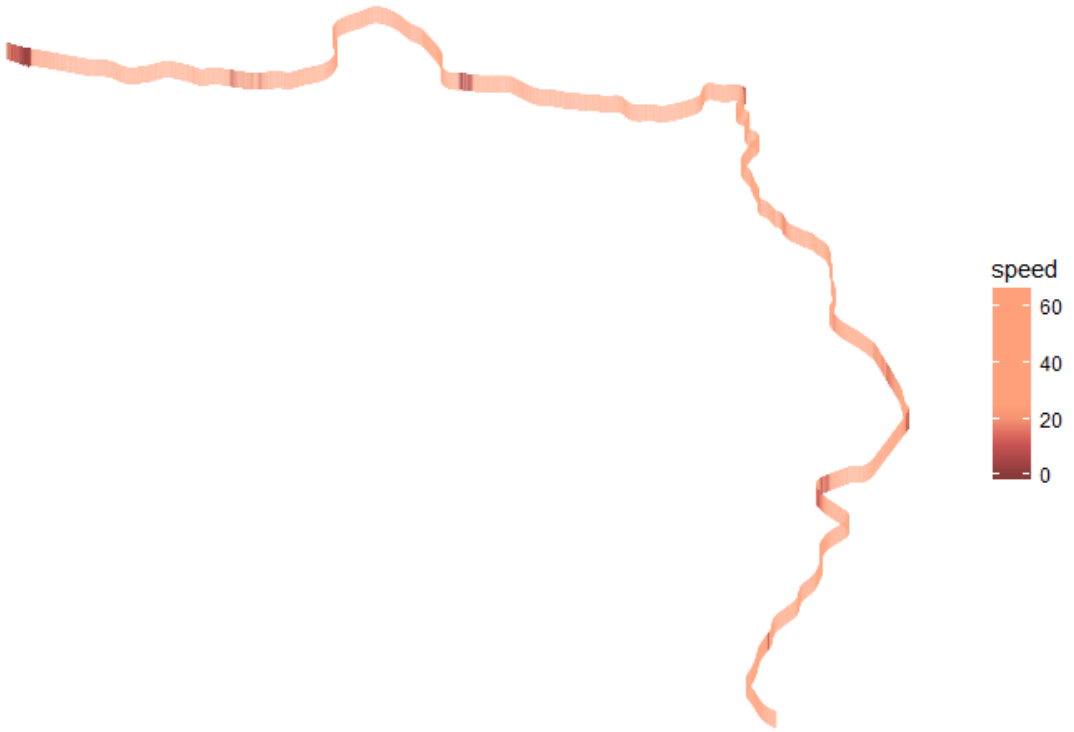
The first figure shows the speeds of all the southbound runs for the peak day along the road. After trying various chart types, it was decided that an error bar chart described the data the best. Each run contains thousands of points, and therefore using lines with no width was the best way since the points would not mask each other. Other types were considered, but it was dificult to see areas were the speed changed just for a few seconds. Also, in order to show all the runs in one graphic, they were all plotted on top of each other with a very low transparency (alpha = 0.05). By keeping all the runs instead of averaging, the figure is truthful (keeps the data as is). This figure in particular is very functional since it shows the speeds at the points on the route that they were recorded, so it better tells the story of where in the path the speeds were slower. The figure is also beautiful since the colors chosen go from a light color (fast speeds) to a dark color (lower speeds). The figure is pleasant to the eye. The figure is also insightful since this graphic includes more than 5 datasets with thousands of rows of data; it would have been very difficult to get anything out of the data without graphic it this way. Lastly, the figure is very enlightening. The reader can quickly see the areas with major slowdowns because it is darker, and the figure can help those making decisions regarding roads in Yellowstone National Park quickly now the problematic areas.

The principles of CRAP were followed by having a defined contrast between the high speeds (light color) and lower speed (very dark color) but also keeping the same "theme" color. This figure in particular only showed the path, and therefore the other principles of CRAP were used in the final graphic.

```
peakday_sb <- bind_rows(all_SB_AM,all_SB_MD)

map_peakday_sb <- ggplot(peakday_sb, aes(x = long, y = lat, color = speed, gr
oup = run)) +
  geom_errorbar(aes(ymin = lat-0.0025,ymax = lat+0.0025), alpha = 0.05) +
  scale_color_gradientn(colours = c("lightsalmon","lightsalmon", "lightsalmon
", "lightsalmon1", "lightsalmon1", "indianred3", "indianred4"),
                        values = seq(60,0,-10)/60) +
  theme_void()

map_peakday_sb
```




```
ggsave(map_peakday_sb, filename = "output/map_SB.pdf", width = 7, height = 5)
```

Figure 2: Southbound Travel Times by Run

The second figure shows the travel time of each southbound run. Because it is important to compare the travel times within each period (AM or PM) and also between periods, one color was used for the same period, i.e., all the runs in the AM were blue and all the runs during midday were green. It was also important to keep the AM runs together as well as the midday run. For this reason, the data was previously divided into a "top", "middle", and "bottom" data frame, with the "top" storing the AM runs, the "middle" a blank row, and the "bottom" the midday runs. The purpose of the "middle" data frame was to have a space that would visually separate the AM runs from the midday runs.

The chart type chosen for this figure was a bar chart. A bar chart was chosen because these runs show a distribution (they are not all the same but they represent the same) and since both periods had to be displayed in the same chart, a bar chart was better than a histogram.

The five qualities of great visualizations were applied in this figure by:

1. Truthful: a chart showing the distribution of all the runs was important to include because it tells the story of how close or far off all the runs are for the same direction and period of time. This gives the reader more context on the data, and makes the overall graphic more truthful. Other than cleaning up the points included in the analysis, the data was not manipulated to show something obscure in the data but rather exactly what the data says.

2. Functional: the graphic is also highly functional because it summarized large datasets into one concise graphic very clearly (each bar represent one dataset/run).

3. Beautiful: the graphic is also beautiful as it uses a nice font and colors to communicate the travel times. The background was also cleaned (no gridlines or colors) to enhance the figure later in Illustrator.

4. Insightful: the raw data contained thousands of rows, which makes the processing of it by the eye almost impossible. Therefore, the figure is very insightful as the reader can quickly see the distribution of travel times for the peak day by period of time.

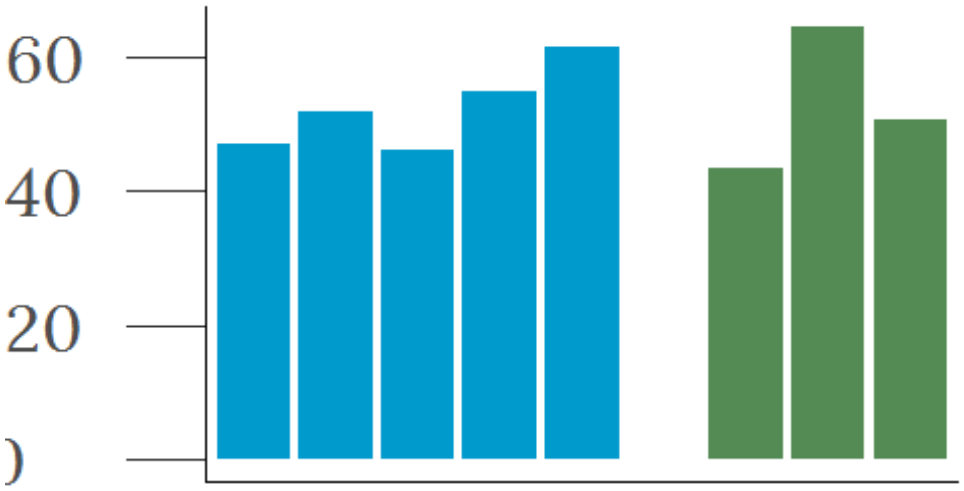
5. Enlightening: the figure is also enlightening as those working in improving the experience of driving this road in the park can see how long it takes to go from the West entrance to Old Faithful, and determine if this is what they expect or if something needs to change to make the experience more pleasant in terms of travel time.

The elements of CRAP were mostly used in the final graphic. In this particular figure, the colors were chosen to have contrast (not too similar, no too different either) and a special font was also used that ties into the overall graphic. All the other elements were applied in the final graphic (since this figure only shows the distribution, no axis titles, titles, etc.).

```
#Load the fonts for the axis text
windowsFonts(lora = windowsFont("Lora"))

#Create a data frame with the plot
sb_runs <- ggplot(data = peakday_sb_tt_v2, mapping = aes(x = run, y = minutes
, fill=period)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("white", "deepskyblue3" ,"palegreen4")) +
  theme_classic(base_family = "lora") +
  theme (axis.text.x=element_blank(),
         axis.ticks.x=element_blank(),
         axis.ticks.length = unit(2, "lines"),
         legend.position="none",
         panel.background = element_blank(),
         plot.background = element_rect(fill = "transparent",colour = NA),
         axis.text.y = element_text(size=25, hjust = -0.6),
         axis.title.y = element_text(size = 25)) +
  labs(x = NULL,
       y = NULL)

sb_runs
```




```
#Save the plot as a pdf
ggsave(sb_runs, filename = "output/SB_runs_chart.pdf", width = 8, height = 5,
bg = "transparent", device = cairo_pdf)
```

Figure 3: Comparison of Travel Times between Peak day and Wild day

The third and last figure shows the comparison of the average travel time of a peak day and the travel time of the wild day by period. The chart chosen for this purpose was a hybrid, the average travel time for the peak day was displayed as a point (the average travel time) with vertical lines (the standard deviation), and the wild day travel times were displayed as horizontal lines. The standard deviation was included in addition to the average travel time because showing this second statistic gives more context to the reader than just showing the average. To show both the average travel time and standard deviation in one symbol, it was decided that a point with vertical lines was the best way to convey that message. The wild day was displayed as horizontal lines to clearly show that they are different datasets.

The five qualities of great visualizations were applied in this figure by:

1. Truthful: the travel times where calculated from roughly the same starting point to the same ending point. Then the average travel time and standard deviation was calculated using all the runs included for the peak day. The wild day also used the same starting and ending point to keep the comparison consistent.

2. Functional: this was accomplished in two ways: (1) using the same type of chart (symbol) for the same day (peak day as a point and vertical lines, and the wild day as horizontal lines), and (2) using the same color for the same period (AM in blue and midday in green). Including the standard deviation in addition to the average also give the reader more context on the data, making this figure more functional.

3. Beautiful: the colors used were chosen to make the figure pleasant to the eye. Also, using different symbols and color to compare the two datasets and periods is in accordance with the purpose of the figure, to compare them.

4. Insightful: it would have been very difficult to see the average travel times and standard deviation from the raw data (thousands of rows and multiple runs). The graphic quickly shows the average of all the runs included as well as the standard deviation so the reader knows how spreadout the data was, and also to compare this with the other periods and days.

5. Enlightening: the figure is also enlightening because it quickly tells the story that the wild day has higher travel times than the average peak day. This helps to those using the graphic to see that the runs were fairly consistent (the standard deviation is small) and the average for the peak day for both the AM and midday periods are very similar. This could really help those making decisions regarding the roads in the park know how long it takes to go through the park, which is very important.

The principles of CRAP were mostly applied in the final graphic, but in this figure the colors were chosen to have contrast. Although not in the same figure, the colors chosen were the same than those chosen in the previous figure.

```
#Create a data frame with the plot
sb_day_comp <- ggplot() +
  geom_pointrange(data =peakday_sb_tt, mapping = aes(x = period, y = ave_tt,
ymin=ave_tt - sd_tt, ymax=ave_tt + sd_tt, color = period),  fatten = 3, size
= 3) +
  scale_color_manual(values = c( "deepskyblue3" ,"palegreen4")) +
  geom_errorbar(data = wildday_sb_tt, mapping = aes(x= period, ymin = minutes
, ymax = minutes, color = period), width = 0.35, size = 3) +
  ylim(0, 80) +
  labs(x = NULL, y = NULL) +
  theme_classic(base_family = "lora") +
  theme(axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.ticks.length = unit(2, "lines"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.background = element_rect(fill = "transparent",colour = NA),
        axis.text.y = element_text(size=20, hjust = -0.6),
        legend.position="none")

sb_day_comp
```


```
#Save the plot as a pdf
ggsave(sb_day_comp, filename = "output/SB_day_comp.pdf", width = 5, height =
5, bg = "transparent", device=cairo_pdf)
```

Final figure

In transportation planning and engineering, almost everything has a spatial component. The project in Yellowstone required knowing two things (1) the speeds at almost every point along the road, and (2) how long it takes from the west entrance to Old Faithful. One of the days in which the data was collected, the driver doing the travel runs experienced an animal traffic jam, and therefore the travel times were much higher. Due to this, it was of interest to compare an average peak day with this wild day. The final figure, therefore, had to have the speeds along the road, the travel times of the runs, and a comparison the travel times during the average peak day and the wild day.

The geometry of the road is horizontal (going from west to east) and then vertical (going from north to south). Due to the geometry of the road, it made sense to have graphic of speeds at various points along the road covering a "corner" of the overall graphic and have everything else inside the void created. The legend of the speed distribution graphic also had to be kept close to the road itself because those two elements of the graphic are related (applying proximity principles). In order to convey the truth and give the reader the context on the project, it was important to include the days data was collected. Therefore, each of the other two charts describe in text when the data was collected and any

information relevant for the results (like what the wild day meant = animal traffic jam). Each of the two graphics were kept in different sections within the overall graphic so the reader understands that the text next to each of the figures corresponds to the one figure and not to the other (apply proximity principle).

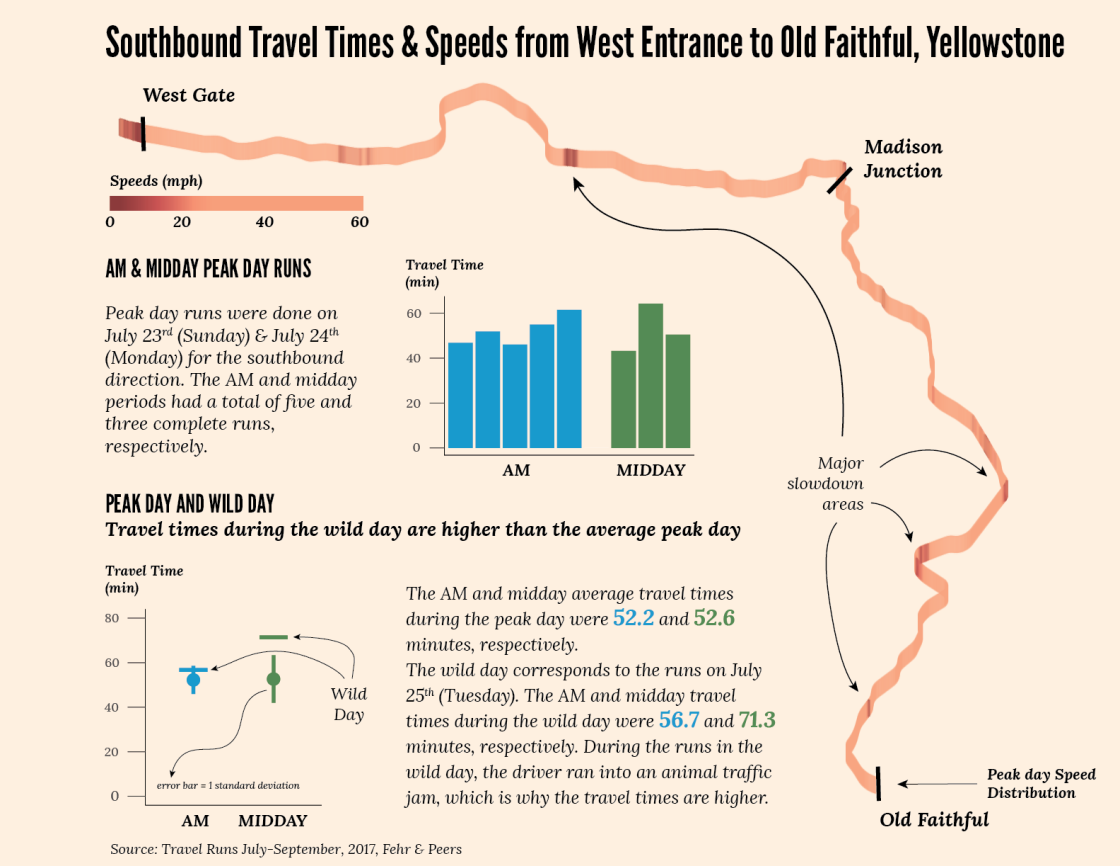
The colors chosen for the graphic were meant to give contrast in the overall visualization but also keep a consistent theme. For the speed distribution graphic, a light coral color was chosen for the higher speeds since people usually relate light colors with good, whereas a darker colors in the same family was chosen to represent slow speeds, since dark usually means bad for people. The two other charts (travel times for each run and the comparison between the peak day and the wild day) were somewhat related. Both were related to the time of day (period). To keep consistency (apply repetition principles), the same color was use to represent the AM period in both graphics, and a different color to represent the midday period in both graphics. For this same reason, in the paragraph explaining the peak day and wild day comparison, the same two colors to represent the AM and midday periods was use to highlight the average travel times for either day (i.e., if the travel time was for the AM peak day or wild day, the color of the value was blue, which is the same color representing the AM runs and the average travel times). Also, blue and green were chosen to represent the AM and midday because gave a good contrast between the background color and the speed distribution graphic.

Fonts was another element that was kept fairly consistent throughout the visualization so the reader feels familiar with graphic. League Gothic was use for the main titles (the title of the graphic and the titles and the two last charts). A very different font (Lora) was used for the subtitle, paragraphs and other design elements to keep contrast. Some text was in bold and italic and other just italics to, again, have contrast and highlight certain things. For example, there are three locations of importance for the reader in the map: the West Gate, Madison Junction and Old Faithful. These three are location in the road and therefore were highlighted with the same font in bold (repetition and contrast). It was also important to highlight that the road shown in the graphic is the speed distribution of the peak day, so this was also in bold. Other elements were highlighted by using arrows, such as showing the areas where lower speeds (major slowdowns) occur in the road. Arrows were used to point to these locations because there were multiple locations and they were spreadout, so by using arrows and describing that those are the major slowdown area, the reader draws the attention to one single point. Arrows were also used to describe the elements in the last chart that correspond to the wild day. It was decided that using the arrows was more effective that having a legend since arrows were used in other areas of the visualization and is more direct.

Alignment was another important element in the visualization. Almost everything in the graphic is left aligned (except for the text for "Major slowdown areas" and "Peak day Speed distribution"). It was decided that it was best to keep everything aligned to the left to keep, again, the reader familiar with the graphic.

The five qualities of great visualizations was also kept in mind while creating the overall graphic. Truth was essential, so the speed distribution shows all the runs on top of each other instead of averaging them. This highlights even more the areas were low speeds

occur. This way the reader knows the areas where speeds are consistently the same (using an average would maybe show areas with a higher speed but in reality some runs had low speeds in that same location, but averaging them does not show it). In addition to displaying the speeds along the road during the peak day, the travel times were calculated using the same two reference points (start and end). This would compare the travel time between the same two points rather than picking two different points for each run. The overall graphic is also highly functional as it tells the whole story. It first shows the speed distribution during the peak day (the graphic draws the attention to that graphic first). The graphic quickly shows the areas were low speeds occur. Next is the distribution of travel times for the AM and midday runs. The colors and location of this graph helps the reader to quickly see that the travel times in both periods are fairly similar, but the midday runs are a little more spreadout than the AM runs. In this context, the reader moves to the next graphic, which shows the average travel times of the peak day compared with the wild day (the day of the animal traffic jam). This last graphic not only shows the average travel time but also the standard deviation that gives more information to the reader. The visualization is also beautiful as it uses a combination of colors and fonts that contrast but also are familiar for the reader. The graphic is also aligned with the purpose, which is to show the speed distribution on a peak day and the travel times from the west entrance to old faithful. The visualization is insightful as it would have been impossible to know where the locations of low speeds are and how long it takes from the west entrance to old faithful by looking at the raw data. The data was massive and the travel runs did not start or end at the same points, so it would have been very hard to find any patterns or understand what the data is telling without processing and making a visualization like the one shown below. The graphic is also very enlightening. This graphic could be used by the decision makers working for park to know which location are experiencing major slowdowns, and how long it takes to go from the west entrance to Old Faithful. This is very important to keep the level of service in the park at a good level and what the graphic shows helps to know the locations for further study.



Results of Southbound Analysis for Yellowston from the West Entrance to Old Faithful