

# Final Project

*Whitney Royal*

10/24/2017

## Executive summary

The question behind this data is how to encourage travel and tourism to Europe. The purpose of the graphics is to enhance marketing data for hotels at Bookings.com and/or travel in general. The final graphic shows the logistics of where to plan a trip to Europe if there are concerns around where to stay. I know this is a personal pain point with traveling, and it is also hard to sort through myriads of data available online through reviews. This graphic provides a quick and easy way to digest the answers to the questions “Should I go to Europe? Where should I go?”

## Data background

The data came from reviews from the Bookings.com website which is available to all. 515,000 customer reviews of 1493 hotels in Europe are in the data along with the geographic location and address of the hotels.

## Data cleaning

Cleaning the data was tricky. The data had to be cleaned to map the coordinates of the hotels onto a map. Additionally, the addresses were sectioned out to identify which country a hotel belonged in. The data also had to be truncated to only include the top reviews since those were the ones that we are interested in and we are trying to solve the pain of having too much, overwhelming data.

## Individual figures

### Figure 1

The first figure was created by cleaning the data and grouping by hotel name. Because each of the reviews had an average score attached, this score did not need to be repeated multiple times. The purpose of this graphic was primarily to obtain a list of the top rated hotels, but the graphic added a greater insight of what the average ranking is. I chose a bar graph since the comparison of the numbers to each other was marginally important and the actual ranking numbers are displayed. I kept the colors aesthetically pleasing and in order with the rankings and made sure the white numbers in the bars provided contrast to make the scores legible. I also kept the image truthful and honest by including tie scores instead of cutting off the data in the middle of a tie.

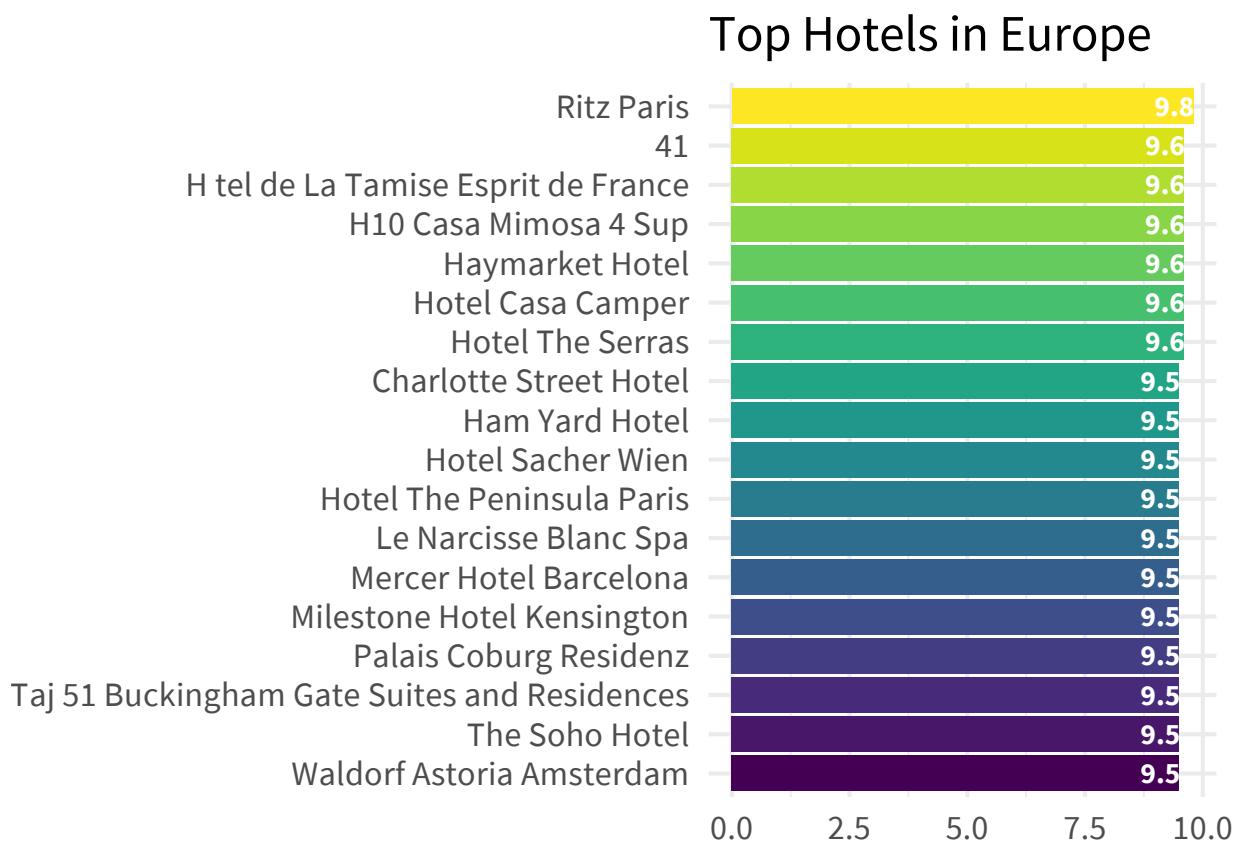
```
newdata <- reviews %>%
  group_by(Hotel_Name) %>%
  slice(1) %>%
  ungroup() %>%
  top_n(10, Average_Score) %>%
```

```

arrange(desc(Average_Score)) %>%
  mutate(Hotel_Name = fct_rev(fct_inorder(Hotel_Name, ordered = TRUE)))

top10 <- ggplot(newdata, aes(x = Hotel_Name, y = Average_Score, fill = Hotel_Name)) +
  geom_col() +
  geom_text(aes(label = Average_Score), nudge_y = -.4, color = "white",
            fontface = "bold", family = "Source Sans Pro") +
  theme_minimal(base_family = "Source Sans Pro", base_size = 16) +
  theme(axis.title.x = NULL, axis.ticks.x = NULL) +
  labs(x = NULL, y = NULL, title = "Top Hotels in Europe") +
  guides(fill = FALSE) +
  coord_flip()
top10

```



**Figure 2**

The second figure was created to show the countries that had the most top hotels. This provides a consumer with a starting point when planning a trip; if there is a country with more options of top hotels, perhaps I choose to travel there so I can have more options. I chose this figure type so the numbers could easily be compared to each other. I kept the countries arranged in alphabetical order since the rank order doesn't seem as important as identifying the countries, and alphabetically arranging the countries felt more digestible than having the numbers in order. I utilized repetition in the colors from the first figure and aligned the country names to the center. I kept the y axis at full length and kept the graph simple and minimal so the clarity of the information could come through. The proximity and contrast were

also considered: originally, I had the lollipops closer together, but the words ran into each other so I spread them apart, and the yellow color was much brighter and matched the first figure better, but the contrast wasn't high enough so I made it less bright. There was definitely a tradeoff happening on both levels: could I keep the color analogous enough to have the feeling of repetition while having a high enough contrast to make it visible?

```
newdata3 <- reviews %>%
  group_by(Hotel_Name) %>%
  slice(1) %>%
  ungroup() %>%
  top_n(100, Average_Score) %>%
  arrange(desc(Average_Score))

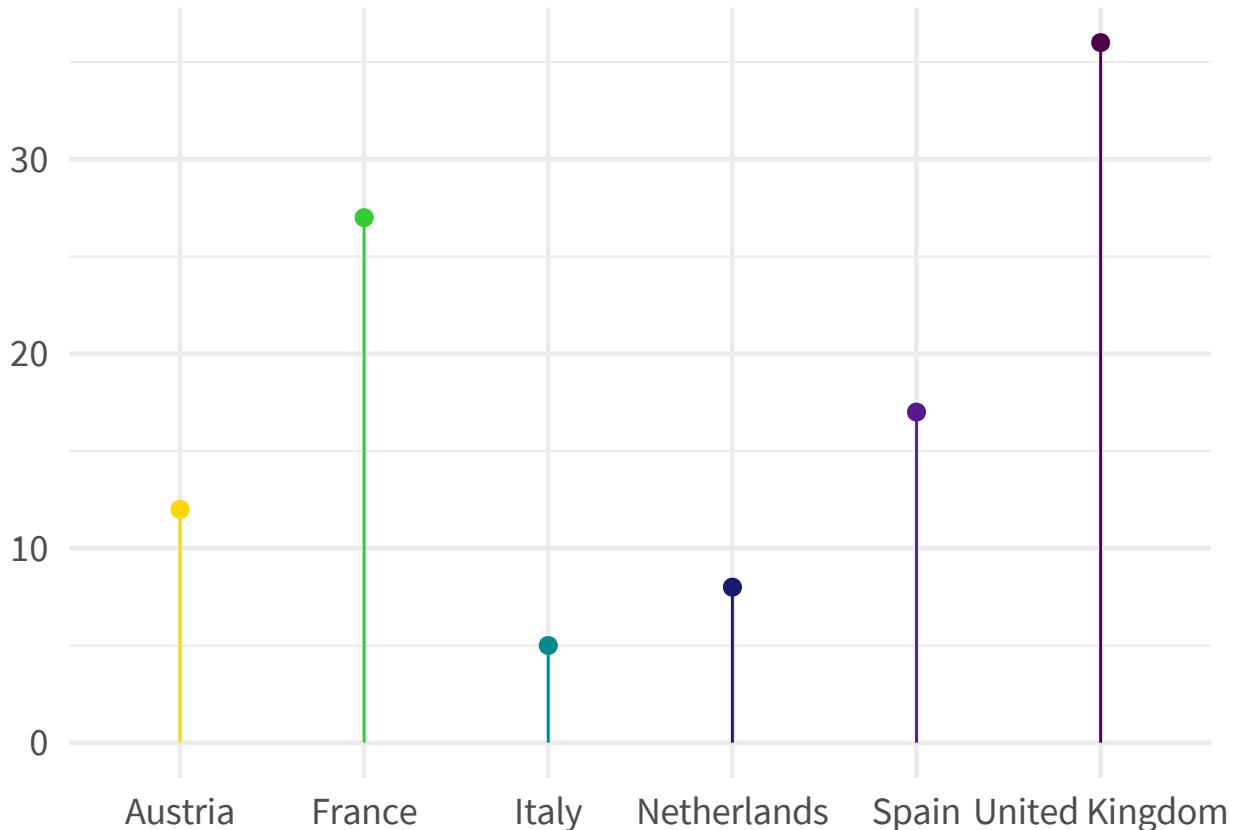
world_shapes <- st_read("data/ne_50m_admin_0_countries/ne_50m_admin_0_countries.shp",
                       stringsAsFactors = FALSE) %>%
  filter(continent == "Europe") %>%
  filter(!(name %in% c("Russia", "Faeroe Is.", "Aland", "Andorra", "Guernsey", "Isle of Man", "Iceland")))

## Reading layer `ne_50m_admin_0_countries` from data source `/Users/whitneyroyal/Desktop/Compa
## Simple feature collection with 241 features and 63 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:            xmin: -180 ymin: -89.99893 xmax: 180 ymax: 83.59961
## epsg (SRID):    4326
## proj4string:    +proj=longlat +datum=WGS84 +no_defs

country_names <- world_shapes %>%
  filter(continent == "Europe") %>%
  distinct(name) %>%
  as_data_frame()

blah <- newdata3 %>%
  regex_left_join(country_names, by = c("Hotel_Address" = "name")) %>%
  group_by(name) %>%
  summarise(n = n())

graphstuffs <- ggplot(data = blah, aes(x = name, y = n)) +
  geom_pointrange(aes(ymin = 0, ymax = n), color = c("gold", "limegreen", "darkcyan",
                                                    "midnightblue", "purple4", "#4d004b")) +
  theme_light() +
  labs(x = NULL, y = NULL) +
  theme(legend.position = "bottom") +
  theme_minimal(base_family = "Source Sans Pro", base_size = 18) +
  theme(axis.title.x = NULL, axis.ticks.x = NULL) +
  labs(x = NULL, y = NULL) +
  guides(fill = FALSE)
graphstuffs
```



**Figure 3**

The third figure is a map with dots to represent where the top 100 hotels are located. Because most of them are in the same cities, the map doesn't differentiate between them all, but it does give an idea of where some good destinations are. I chose this map because having a visual interpretation of the locations is liked by people when traveling. It allows them to see where they would be going and makes the experience seem more tangible. The map wouldn't be used for accuracy but is more of a frame of reference for the traveler. Because it is just for perspective, I took the contrast and proximity into account. I chose the aesthetics to be simple since the map doesn't have the most important information - it's mostly a nice add-on that sells the experience to the consumer. It also truthfully and functionally shares this information and enlightens travelers to the possibility of visiting this continent. It also reveals the locations in proximity to other countries (and their own countries) in a way that is more difficult to see with addresses alone.

```
dataformaps <- newdata3 %>%
  st_as_sf(coords = c("lng", "lat"), crs = 4326) %>%
  st_transform(crs = 102014)
world_shapes2 <- world_shapes %>%
  filter(!(name %in% c("Russia", "Faeroe Is.", "Aland", "Andora", "Guernsey", "Isle of Man", "."))
pointsonly <- st_coordinates(dataformaps)

dataformapsall <- cbind(dataformaps, pointsonly)

#if it works in google, it's 4326
themap <- world_shapes
```

```

mapofhotels <- ggplot() +
  geom_sf(data = world_shapes2, color = "gray") +
  geom_point(data = dataformapsall, aes(x = X, y = Y), color = "limegreen", size = 2) +
  coord_sf(crs = st_crs(102014)) +
  theme_void() +
  theme(legend.position = "bottom")

mapofhotels

```

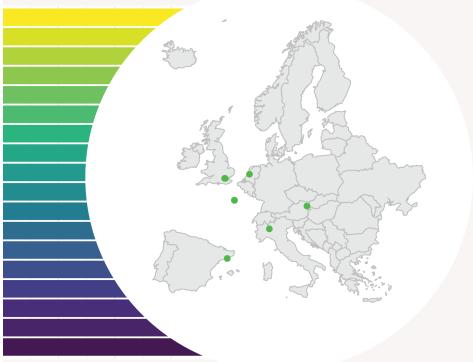


## Final figure

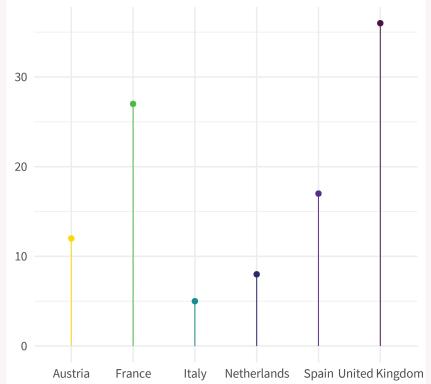
I designed this piece to be appealing and enticing to a traveler. I wanted it to be simple and engaging while having a pop of fun and color. I chose the colors because they draw attention to the data and provide a cohesive and enticing landscape for travel. They also echo the brightly colored buildings of the images. I repeated the same gradient of the graphic on the opposite side of the final figure to bring in continuity and balance. I chose the font Source Sans Pro because it is my new favorite font: it is clean and simple and awesome. I added images that inspire travel to Europe to invite the observer in and to bring life to the data. It conveys truth of the top options in Europe and allows the reader to feel confident in choosing to visit due to the prevalence of highly rated hotels. I kept the data for the other 1300 hotels out of the visualizations since too much data is overwhelming and since we only care about visiting the best. I imagine this image being used in a flyer, pamphlet, or print medium such as a magazine (the proportions even leave space for a bar of advertising on the side - spoken like a true business student).

# Visit Europe

Looking to visit Europe but don't know where to stay?  
We've identified top hotels in Europe based on over  
a half million Bookings.com reviews.



Number of top hotels per country



Top hotels in Europe

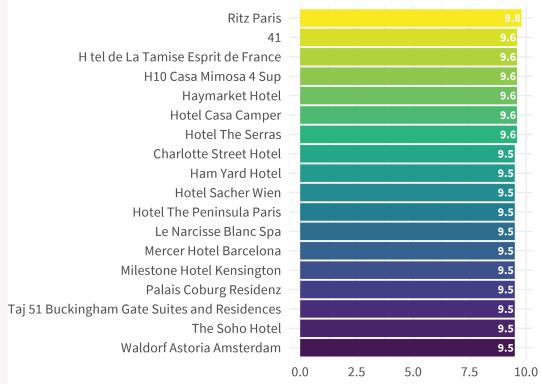


Figure 1: This is where the magic happens