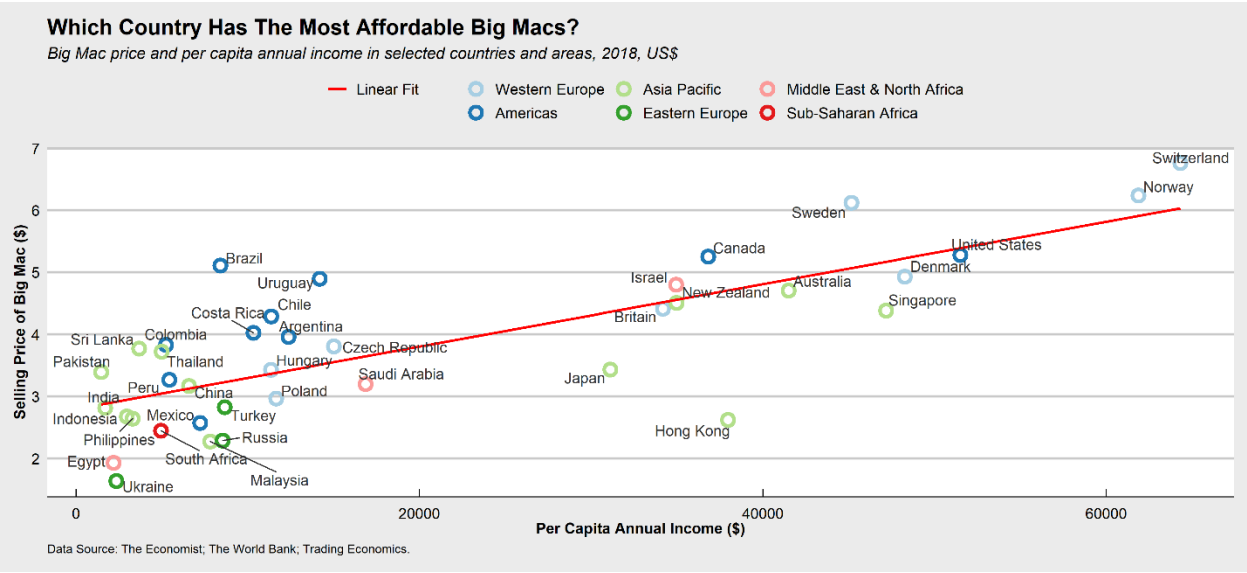
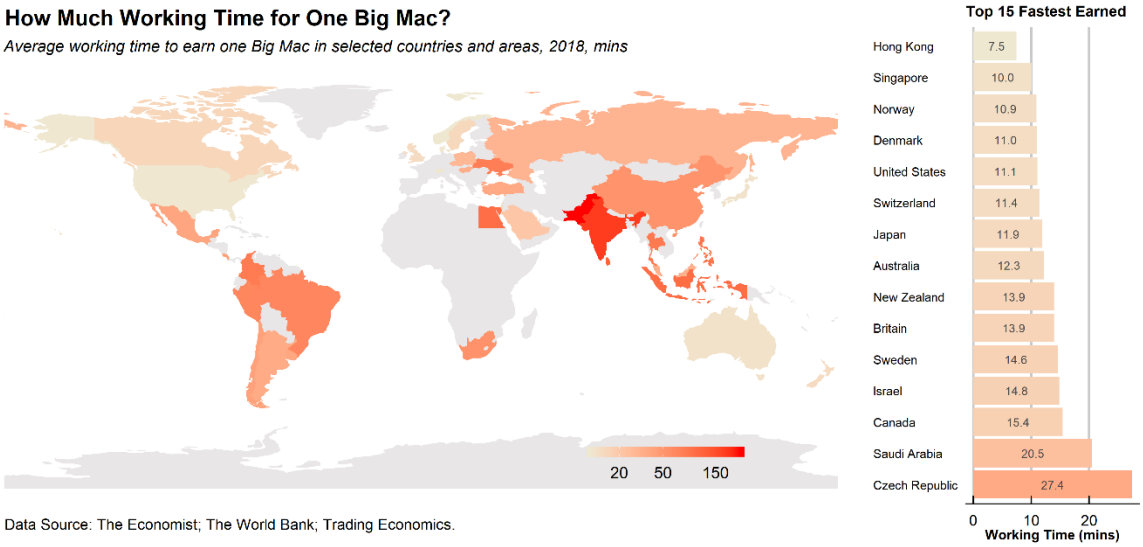
In this assignment, I choose the Big Mac Index dataset for both of my visualizations. Originally, the Big Mac Index was designed to demonstrate the "correctness" of currencies' exchange rates. However, I personally care more about the Big Mac itself rather than its importance in the finance world. As a Big Mac lover, I am interested in finding the country with the most affordable big macs, in case I get unemployed after my graduation (hopefully this won't happen). To successfully deliver this quantitative message, I merge the original dataset with the income per capita dataset from the World Bank. In addition, I use the data from Trading Economics to fix the NA value for Hong Kong. Since there are lots of countries still missing income data for 2019, I choose January 2018 to be the time for both the Big Mac price data and the income per capita data for uniformity.

For the first visualization, I decide to use a scatter plot with a linear regression line, so that it's possible to quickly tell whether Big Mac is affordable in a certain country. Those points above the linear fit line are countries with over-priced Big Macs (compared to their per capita income), while the points below the linear fit line are countries with affordable Big Macs. In this visualization, I use geom_point(shape=1,size=3,stroke=2) for the points to make it easier to tell in overlapping situations (e.g. Indonesia and Philippines). In addition, I color the points by the region of each country, so the readers can get a quick grasp of the distribution for countries in different regions. For the linear fit line, I uses geom_smooth() and specify linetype=paste("Linear Fit") inside the aes(). Combining this with specifying scale_linetype(name=NULL), I manage to remove the word "linetype" from the legend, which is redundant information in this case. After that, I use geom_text_repel() to label the points with corresponding country names. I use a color brewer with a "Paired" palette to enable readers to easily tell the difference between different regions. In this visualization, I use the white economist theme from the "ggthemes" package to get the classic Economist look. Adding on that, I specify the font of title to be bold and larger and italicize the subtitle to make the difference between these two more obvious. Finally, I call the labs() to add a footnote of data source for completeness. With these elements, this visualization enables readers to understand the distribution of countries from different regions in terms of Big Mac price and per capita income while quickly categorizing them into "more affordable" and "less affordable" groups.



In the second graph, I try to find out how many minutes an average worker has to work to earn enough money for a Big Mac in different countries. To intuitively show the distribution around the globe, I choose geospatial visualization as primary tool, while using a bar plot to show the top 15 fastest earned countries, which I care very much as discussed previously. I use 4 additional packages to plot the geospatial plot, “sf”, ”rnaturalearth”, ”rnaturalearthdata”, and “rgeos”. I first call ne_countries() to get a data frame of countries and their geospatial data. Then, I merge the dataset with the Big Mac dataset and the income per capita dataset by the 3-letter country code. After that, I calculate the working time to earn a Big Mac by using the formula – time = dollar_price / (annual income per capita / working hours) * 60, assuming annual working hours to be 1800 hours for all countries. I make this assumption due to the reason that working hours data are very hard to find, and I will generate NA for almost 50% of the countries even with the most complete dataset. I specify scale=”small” here, because the small scale excludes small islands from the graph, making it visually clean and less distracting. When plotting the world map, I use two layers of geom_sf(), one with a uniform fill of light gray and one being colored by actual working time data. I call the scale_fill_gradient() function to set a continuous color scale for the world heat map. In particular, I choose Beige as the “low” color and Red as the “high” color to create a harmonic palette. To make the color difference even easier to spot, I specify trans=”log” in the scale_fill_gradient() function, so that countries with low working time can be compared. I also specify breaks to be c(20,50,150) to avoid float-type color bar tick values. I use theme_map() for the geospatial part to get a clean and minimal white background with no grid line or axis tick. In the theme settings, I assert the legend.position to be c(0.67,0.1). This places the legend on the Antarctica of the map, which doesn’t have a working time value. For the bar plot part, I only use the top 15 countries with least working time. For the continuous color bar, I use Beige for the “low” color, while choosing a specific “#FFAA85” for the “high” color. This is because if I still use red for the “high” color, the color scale of the bar plot won’t match the color scale of the world map. For this part, I also use the economist white theme, but I set plot.background to be white in the theme setting to get a white background like what the world map has. Just like what I did in the previous visualization, I label the title, subtitle, and footnote for both parts. Finally, I utilize the ggdraw() and draw_plot() functions from the “cowplot” package to combine the two parts together. From this visualization, readers should easily find out the best location for an low-income Big Mac lover is Hong Kong, where people on average only work for 7.5 mins to earn a Big Mac.



Additional Dataset Used

The World Bank (2018). *Adjusted Net National Income Per Capita.* https://data.worldbank.org/indicator/NY.ADJ.NNTY.PC.CD

Trading Economics (2018). *Hong Kong Income Per Capita.* https://tradingeconomics.com/hong-kong/indicators