Visualization Challenge

Jannik Paul

2022-06-24

Contents

Guidelines

- Plots are supposed to be
  1. interesting
  2. effective (we humans can grasp the message)
  3. scientifically appropriate (don't mislead)
  4. aesthetically pleasing.
- Explanations are supposed to be in full sentences (not only bullet points)

The data

This challenge uses data from https://happyplanetindex.org. On that website a community of international organisations and indidviduals publish an index - the Happy Planet Index - which is meant as an alternative to mainstream indicators of economic growth.

As stated on the website: *"Our current economic system is driven by a 'growth at all costs' mentality, as measured by Gross Domestic Product (GDP). There is an entrenched belief that GDP growth is synonymous with increasing well-being and prosperity and is universally beneficial. In reality, GDP growth on its own does not mean a better life for everyone, particularly in countries that are already wealthy. It doesn't take into account inequality, the things that really matter to people like social relations, health, or how they spend their free time, and crucially, the planetary limits we are up against."*

The Happy Planet Index (HPI) is computed from three variables: (1) life expectancy, (2) experienced well-being, (3) ecological footprint. The first two contribute positively to the index, the third one contributes negatively to the index. You can find the details here, in particular a precise definition of the variables (page 3 and 4). The data set also contains the variable GDP per capita, although it is not used for the computation of the Happy Planet Index.

```r
# Load your packages here
library(readr)
library(tidyverse)
library(ggplot2)
library(hrbrthemes)
library(ggrepel)
library(reshape2)
library(sf)
library(rnaturalearth)
library(rgeos)
library(scico)
library(ggridges)
```

```r
# Read in the data set here
hpi <- read_csv("happy-planet-index.csv")
show(hpi)
```

```
## # A tibble: 1,964 x 10
##    iso   country      continent    year   hpi lifeexp experienced_wel~ footprint
##    <chr> <chr>        <chr>       <dbl> <dbl>   <dbl>            <dbl>     <dbl>
##  1 GTM   Guatemala    Latin Amer~  2006  53.1    70.1             5.90     1.75
##  2 SAU   Saudi Arabia Middle Eas~  2010  39.8    73.9             6.31     5.65
##  3 NLD   Netherlands  Western Eu~  2020  56.8    81.6             7.50     4.21
##  4 SRB   Serbia       Eastern Eu~  2020  53.5    75.7             6.04     2.37
##  5 CRI   Costa Rica   Latin Amer~  2007  63.9    78.4             7.43     2.76
##  6 POL   Poland       Eastern Eu~  2018  44.3    78.5             6.11     4.76
##  7 ZAF   South Africa Africa       2020  34.3    60.6             4.95     2.73
##  8 MOZ   Mozambique   Africa       2016  38.6    58.3             4.41     0.771
##  9 TZA   Tanzania     Africa       2015  34.5    63.1             3.66     1.35
## 10 MNG   Mongolia     East Asia    2012  28.2    68.2             4.89     6.35
## # ... with 1,954 more rows, and 2 more variables: gdp_capita <dbl>,
## #   biocapacity <dbl>
```

```r
# hpi is the full unaltered data frame (encompassing all years before the reduction to 2019)
```

Task 1 (30 points)

- Focus only on the year 2019.
- The task is to visually reveal relationships between multiple variables: How are `gdp_capita`, `experienced_wellbeing`, `footprint` and `life_expectancy` related to the Happy Planet Index `hpi`? And how are these variables interrelated?
- Create a maximum of 2 plots to expose the most interesting types of relationships. The plot should also contain (a selection of) country names.
- Provide a brief textual explanation of your main insights.

```r
# Focus only on the year 2019 and filter the data
df <- hpi %>% filter(year == 2019)
# Take a look at the top countries arranged by hpi
head(df %>% arrange(desc(hpi)), 10)
```

```
## # A tibble: 10 x 10
##    iso   country     continent     year   hpi lifeexp experienced_wel~ footprint
##    <chr> <chr>       <chr>        <dbl> <dbl>   <dbl>            <dbl>     <dbl>
##  1 CRI   Costa Rica  Latin Ameri~  2019  62.1    80.3             7.00      2.65
##  2 VUT   Vanuatu     East Asia     2019  60.4    70.5             6.96      1.62
```

```
## 3 COL   Colombia    Latin Ameri~  2019  60.2   77.3             6.35      1.90
## 4 CHE   Switzerland Western Eur~  2019  60.1   83.8             7.69      4.14
## 5 ECU   Ecuador     Latin Ameri~  2019  58.8   77               5.81      1.51
## 6 PAN   Panama      Latin Ameri~  2019  57.9   78.5             6.09      2.10
## 7 JAM   Jamaica     Latin Ameri~  2019  57.9   74.5             6.31      1.84
## 8 GTM   Guatemala   Latin Ameri~  2019  57.9   74.3             6.26      1.77
## 9 HND   Honduras    Latin Ameri~  2019  57.7   75.3             5.93      1.58
## 10 URY  Uruguay     Latin Ameri~  2019  57.5   77.9             6.60      2.62
## # ... with 2 more variables: gdp_capita <dbl>, biocapacity <dbl>
```

These are the 10 happiest countries according to the HPI.

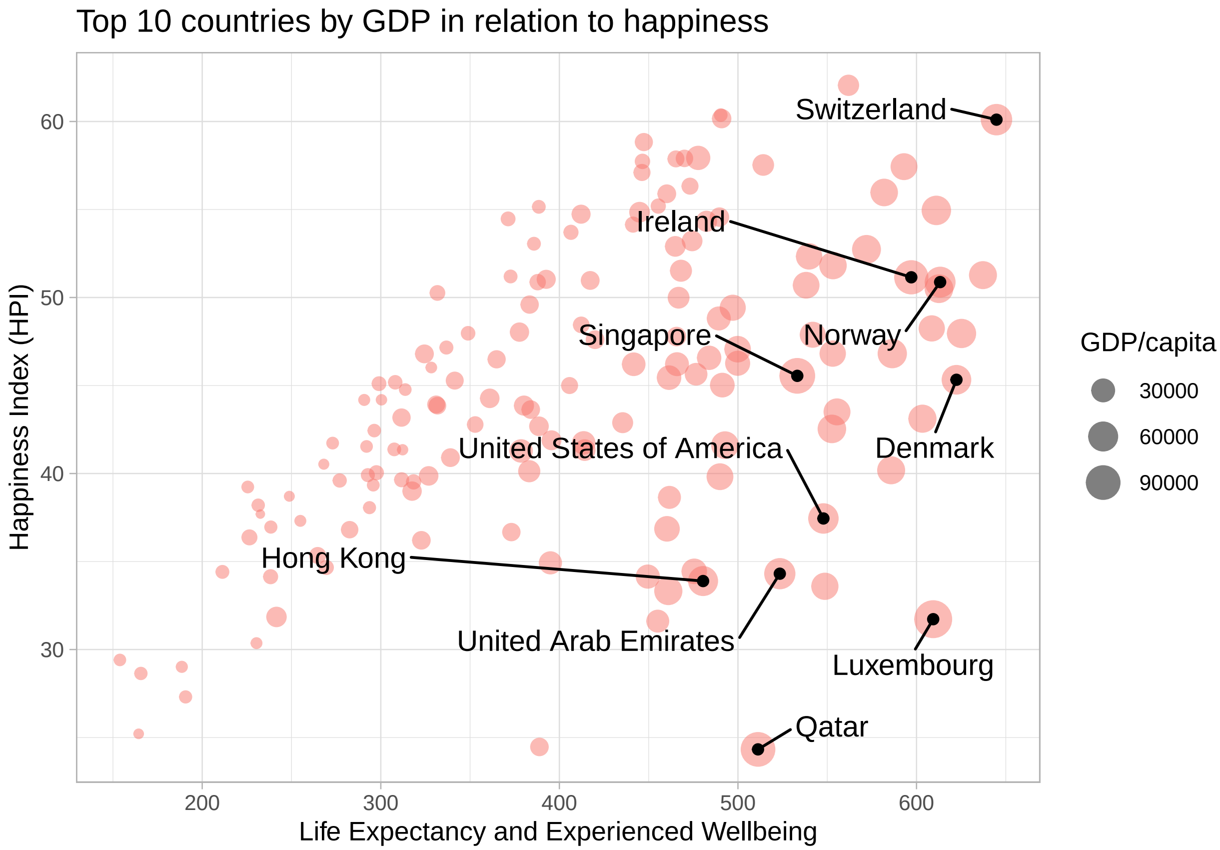
Labeling 152 countries in the plot would overwhelm the recipient. We want to make it visually understandable and pleasing.

```
# Happiest 10
df_hpi_top10 <- df %>%
  slice_max(hpi, n=10)

# Unhappiest 10
df_hpi_bottom10 <- df %>%
  slice_min(hpi, n=10)

# 10 Largest ecological footprints
df_footprint_top10 <- df %>%
  slice_max(footprint, n=10)

# Happiest and unhappiest 10
df_hpi_top_bottom10 <- df %>%
  slice_min(hpi, n=10) %>%
  bind_rows(df %>%
     slice_max(hpi, n=10))

# 10 largest gdp per capita
df_richest_10 <- df %>%
  slice_max(gdp_capita, n=10)
```

First Plot

> We know that life expectancy and experienced well being are scaled in the formula the HPI is calculated from. They both are adjusted for inequality which is unfortunately not represented in the data. They are scaled and multiply to increase the HPI, so we combine them. I want to show how the GDP really does not play a role in calculating happiness for the HPI in this model.

```
ggplot(data = df, mapping = aes(x=lifeexp*experienced_wellbeing,y=hpi))+
  geom_point(aes(size=gdp_capita, color="grey"), alpha=0.5) +
  guides(colour = "none") +
  geom_point(data=df_richest_10, color="black") +
  geom_text_repel(data=df_richest_10,box.padding = 1.2, max.overlaps = Inf, aes(label=country)) +
  guides(size = guide_legend("GDP/capita")) +
  theme_light(base_size = 10) +
  xlab("Life Expectancy and Experienced Wellbeing") + ylab("Happiness Index (HPI)") +
  ggtitle("Top 10 countries by GDP in relation to happiness")
```

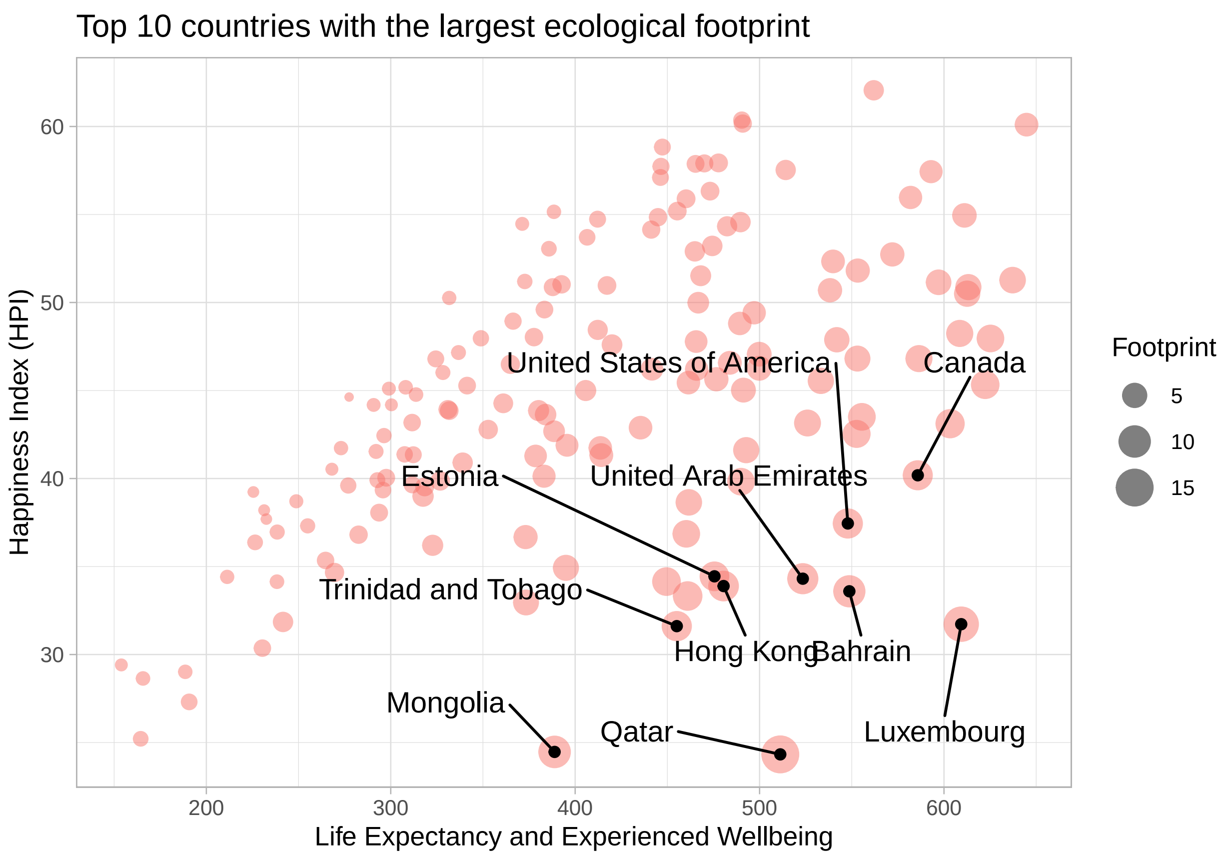


> We see that the countries of high life expectancy and well being usually have a higher than average GDP. We also see that the countries with highest GDP per capita are not the happiest in this model. The span of GDP is all over the place, from the happiest to the unhappiest!
This proves what is mentioned in the documentation - GDP does not affect happiness, and we have just shown this. But what makes countries happier? *There must be other factors playing a part here.*

Second Plot

The economical footprint affects HPI by setting the divider - the larger it is, the smaller our HPI will become. We want to display whether this is true, and show the recipient the pull downwards to unhappiness in countries that have a high ecological footprint. It is the average space a person requires in this country to live unaffected in their typical consumption (*not production*) pattern.

```
ggplot(data = df, mapping = aes(x=lifeexp*experienced_wellbeing,y=hpi))+
  geom_point(aes(size=footprint, color="grey"), alpha=0.5) +
  guides(colour = "none") +
  geom_point(data=df_footprint_top10, color="black") +
  geom_text_repel(data=df_footprint_top10,box.padding = 1.2, max.overlaps = Inf, aes(label=country)) +
  guides(size = guide_legend("Footprint")) +
  theme_light(base_size = 10) +
  xlab("Life Expectancy and Experienced Wellbeing") + ylab("Happiness Index (HPI)") +
  ggtitle("Top 10 countries with the largest ecological footprint")
```



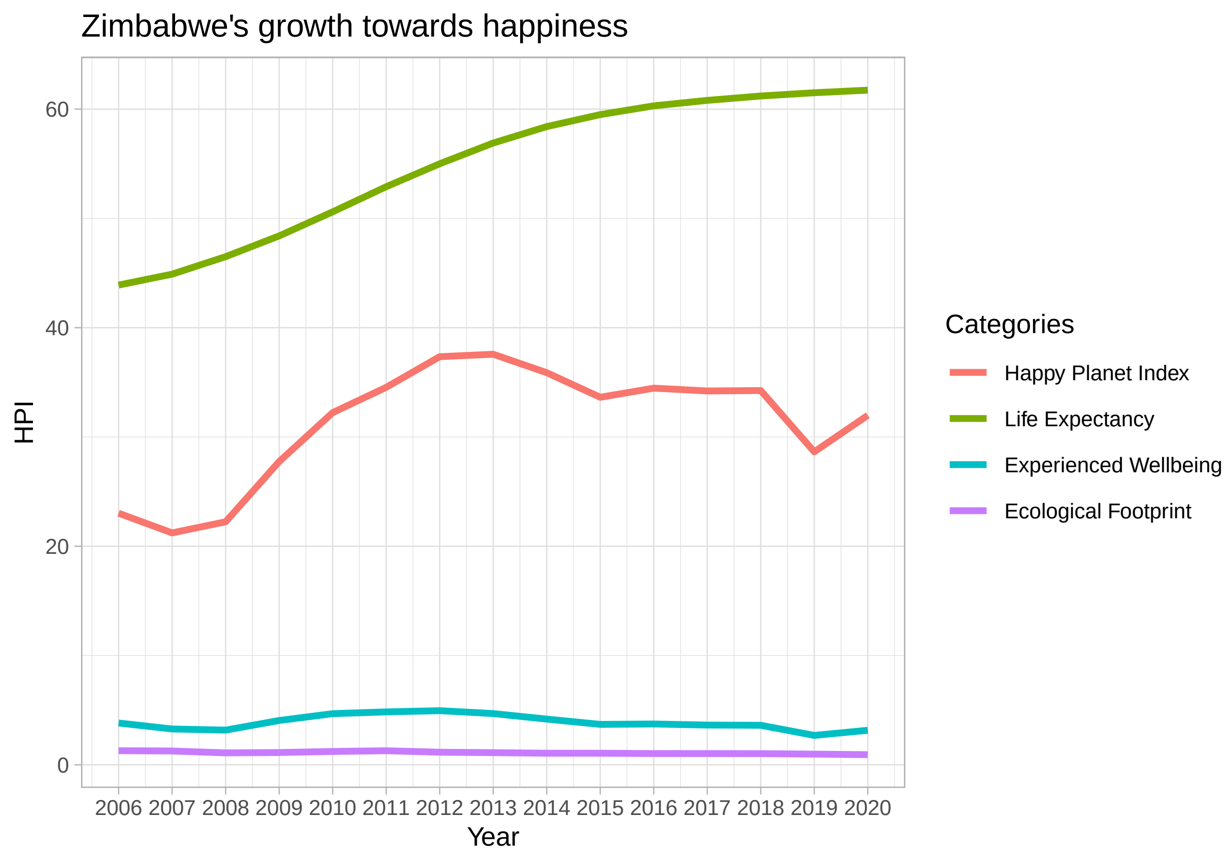
> We see that the larger the footprint, the larger the pull away from happiness is. Man seems to like to live in balance with nature. Both Mongolia and Qatar are incredibly hostile places to man, mostly deserts none could survive in nowadays without clinging to their metropolitan areas.
We have thus shown that the larger ecological footprints reduce happiness in the HPI model. *This echoes in the formula used to calculate the HPI!*

Task 2 (30 points)

- Focus only on the country Zimbabwe.
- The task is (1) to to show how Zimbabwe's Happy Planet Index has evolved in the course of time, and (2) to expose the reasons for this change.
- Create a single plot that contains the time series of `lifeexp`, `experienced_wellbeing`, `footprint`, and `hpi` between 2006 and 2020.
- Provide a brief textual explanation of your main insights.

```
df_zimbabwe <- hpi %>%
  filter(country == "Zimbabwe") %>%
  select(-c(biocapacity,iso,country,continent,gdp_capita))

# I am going to melt the dataframe and I want the variables to be expressive to read
colnames(df_zimbabwe) <- c("Year","Happy Planet Index","Life Expectancy","Experienced Wellbeing","Ecolog

# Melting the dataframe, an interesting technique that even tackles NaNs properly
m_df_zimbabwe <- melt(df_zimbabwe,id="Year")

# Prepare column names so I don't need to haggle too much with labels later
```

```r
colnames(m_df_zimbabwe) <- c("Year","Category","Value")

m_df_zimbabwe %>%
  ggplot(mapping = aes(x=Year,y=Value,colour=Category)) +
  geom_line(size = 1.2) +
  theme_light(base_size = 10) +
  xlab("Year") + ylab("HPI") +
  #theme(axis.text.y = element_text(colour="white", size=0)) +
  ggtitle("Zimbabwe's growth towards happiness") +
  guides(col=guide_legend("Categories")) +
  scale_x_continuous(breaks = round(seq(min(m_df_zimbabwe$Year), max(m_df_zimbabwe$Year), by = 1),1))
```

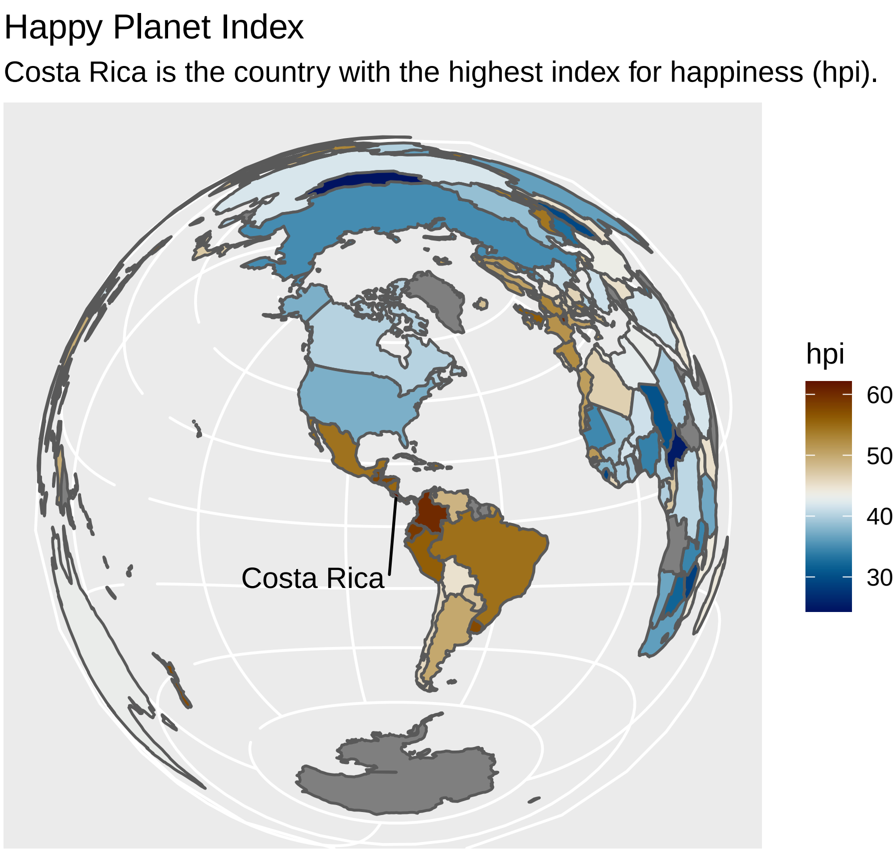


> We see that the *Happy Planet Index* (HPI) is growing steadily. We can see that both *Experienced Wellbeing* and the *Ecological Footprint* stayed on a similar level. *Life Expectancy* rose from 2007 onwards and pushed the *HPI* along with it. When *Life Expectancy* started leveling off to drop in 2015, the growth of *HPI* also lost momentum and started leveling off. With the sharp dip of *Life Expectancy* due to the hunger crisis from lack of rain in 2019 (Link: https://news.un.org/en/story/2019/12/1052621), Zimbabwe's *HPI* growth will stall further or may even sink again.

Task 3 (25 points)

- Visualize the Happy Planet Index of the year 2019 on a map of the world
- Choose a LAEA projection of the world (i.e. the world gets represented as a globe), and put the country with the highest Happy Planet Index in the center of the world.

```r
# Read in simple feature data for our maps
geo <- rnaturalearth::ne_countries(returnclass = 'sf') %>% select(iso_a3, geometry, continent, name_long

# Join data frame with sf object
sf <- df %>% right_join(geo, by = c("iso" = "iso_a3")) %>%
  sf::st_as_sf()

# Prepare our base plot that we want to display
map <- sf %>%
  ggplot(aes(fill=hpi)) +
  geom_sf(aes()) +
  scico::scale_fill_scico(palette = "vik")

# Slice the country highest in HPI (Costa Rica) so we can use it for our repelled text
costarica <- df %>% slice_max(hpi)

map +
  coord_sf(crs = st_crs("+proj=laea + lat_0=9.934739 + lon_0=-84.087502")) +
  geom_text_repel(x=0, y=0, data=costarica, box.padding = 1.8, max.overlaps = Inf,
                  aes(label="Costa Rica")) +
  labs(title = "Happy Planet Index",
       subtitle = "Costa Rica is the country with the highest index for happiness (hpi).")
```

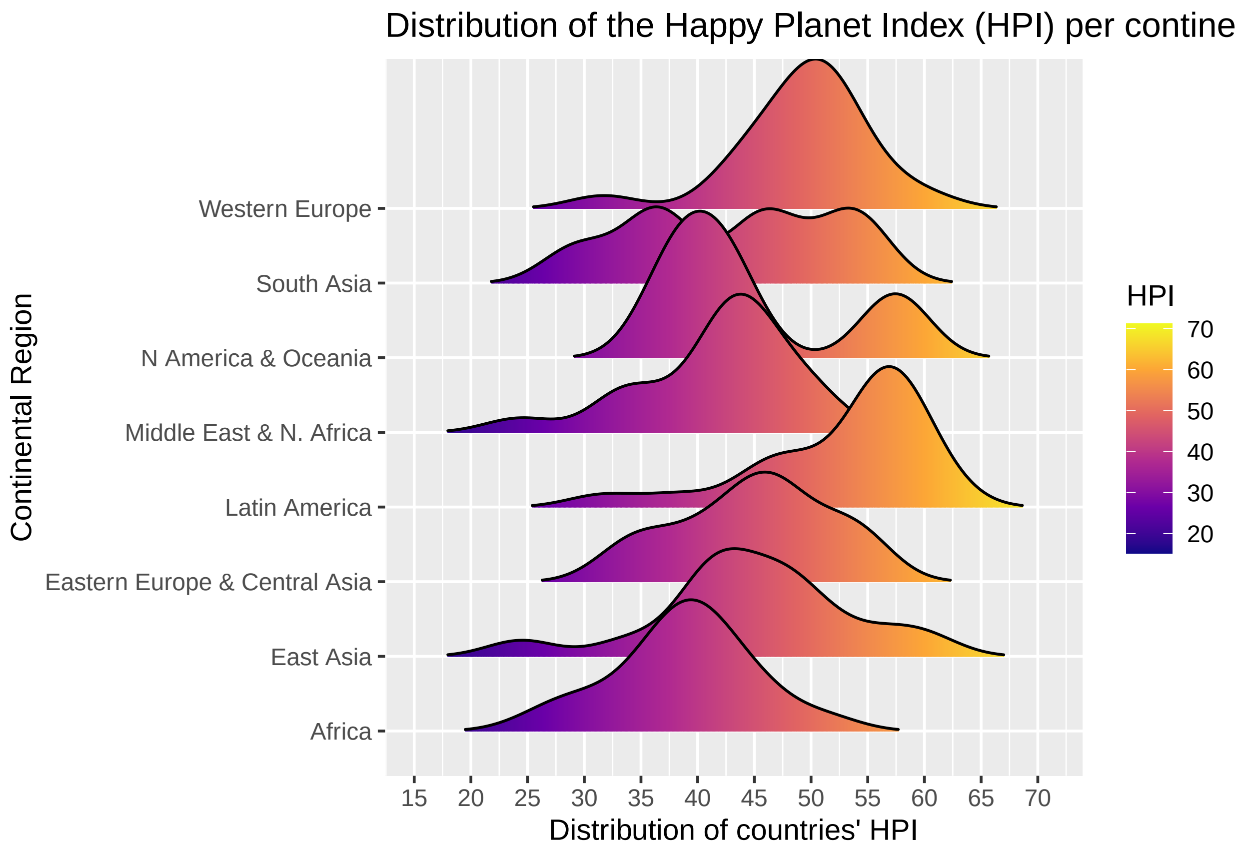


> We know from the Exploration that Costa Rica enjoys the highest HPI. I have a friend from Costa Rica living near Seattle, he would sign this assertion immediately. The map focuses on Costa Rica and the happiness index is portrayed on an expressive scale from cold to hot.

Task 4 (15 points)

- Create a plot that shows the distribution of the Happy Planet Index per continent. While there are many possible ways to represent distributions, choose one that makes it easy to compare the different continents with each other.

```
# This may be my most elegant plot yet
# https://cran.r-project.org/web/packages/ggridges/vignettes/introduction.html
# https://r-charts.com/distribution/ggridges/

ggplot(df, aes(x = hpi, y = continent, fill = stat(x))) +
  geom_density_ridges_gradient(scale = 2, rel_min_height = 0.01) +
  scale_fill_viridis_c(name = "HPI", option = "C") +
  scale_x_continuous(breaks = round(seq(min(15), max(80), by = 5),1)) +
  xlab("Distribution of countries' HPI") + ylab("Continental Region") +
  ggtitle("Distribution of the Happy Planet Index (HPI) per continent")
```



> This plot does not just show the absolute aggregated values of the *continents*, but also the distribution of the *countries* within them, while it is *scaled automatically* so that no continent with many countries dominates. We see immediately that *Western Europe* has a decent amount of countries with relatively high HPIs concentrated around 52. Hence we can also see something interesting: The *equality in HPI by country* within Western Europe is pretty good. Latin America however shows to have a very *high* leaning distribution of the Happy Planet Index and marks the top of the world by this standard.