

# Happiness around the World (2015)

Josh Stremmel, Jooan Choi, Namrata Singh

## Table of contents

Overview . . . . .	1
Data Import . . . . .	1
Exploratory Plot . . . . .	2
Data Tidying and Cleaning . . . . .	2
Exploratory Data Analysis . . . . .	4
Appendix . . . . .	5

## Overview

The World Happiness Report puts a number on something we all care about but rarely measure: how happy people are, country by country. Our project digs into the 2015 report to see what really drives those scores— is it money, health, freedom, or something else? By playing with the data we want to tell a short, visual story about what separates the happiest places from the not-so-happy ones, and maybe learn what our own country could do better. It's basically a data-powered vibe check for the whole planet.

## Data Import

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.4
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

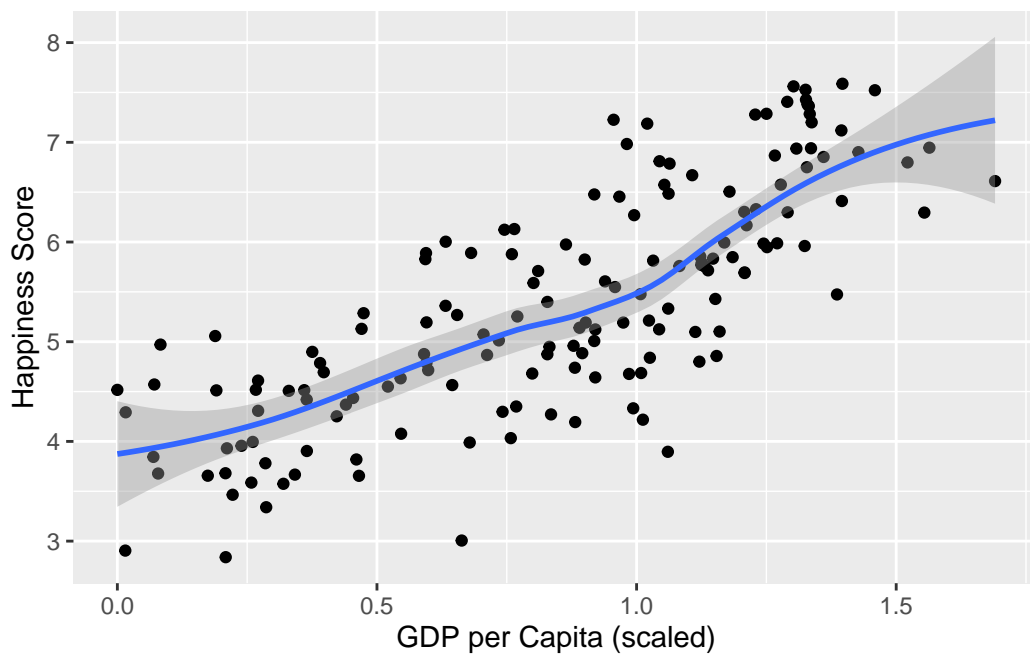
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

## Exploratory Plot

``geom_smooth()`` using method = 'loess' and formula = 'y ~ x'



## Data Tidying and Cleaning

Table 1: Data summary

Name	happy
Number of rows	158
Number of columns	12
Column type frequency:	

character	1
factor	1
numeric	10
<hr/>	
Group variables	None
<hr/>	

#### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
country	0	1	4	24	0	158	0

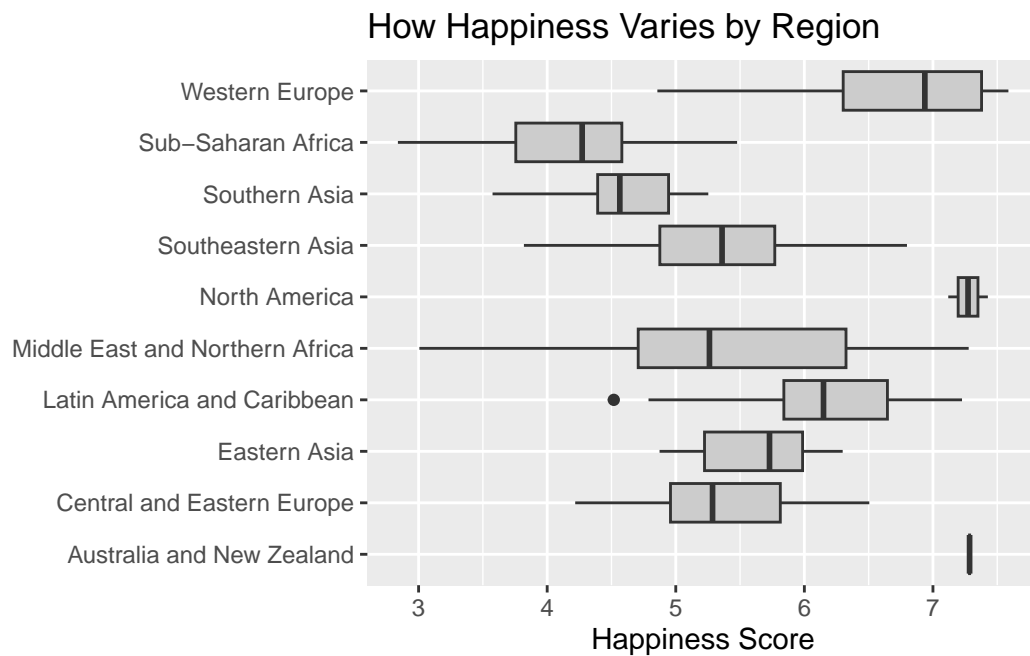
#### Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
region	0	1	FALSE	10	Sub: 40, Cen: 29, Lat: 22, Wes: 21

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
happiness_rank	0	1	79.49	45.75	1.00	40.25	79.50	118.75	158.00	
happiness_score	0	1	5.38	1.15	2.84	4.53	5.23	6.24	7.59	
standard_error	0	1	0.05	0.02	0.02	0.04	0.04	0.05	0.14	
economy_gdp_per_capita	0	1	0.85	0.40	0.00	0.55	0.91	1.16	1.69	
family	0	1	0.99	0.27	0.00	0.86	1.03	1.21	1.40	
health_life_expectancy	0	1	0.63	0.25	0.00	0.44	0.70	0.81	1.03	
freedom	0	1	0.43	0.15	0.00	0.33	0.44	0.55	0.67	
trust_government_corruption	0	1	0.14	0.12	0.00	0.06	0.11	0.18	0.55	
generosity	0	1	0.24	0.13	0.00	0.15	0.22	0.31	0.80	
dystopia_residual	0	1	2.10	0.55	0.33	1.76	2.10	2.46	3.60	

## Exploratory Data Analysis



Registered S3 method overwritten by 'GGally':

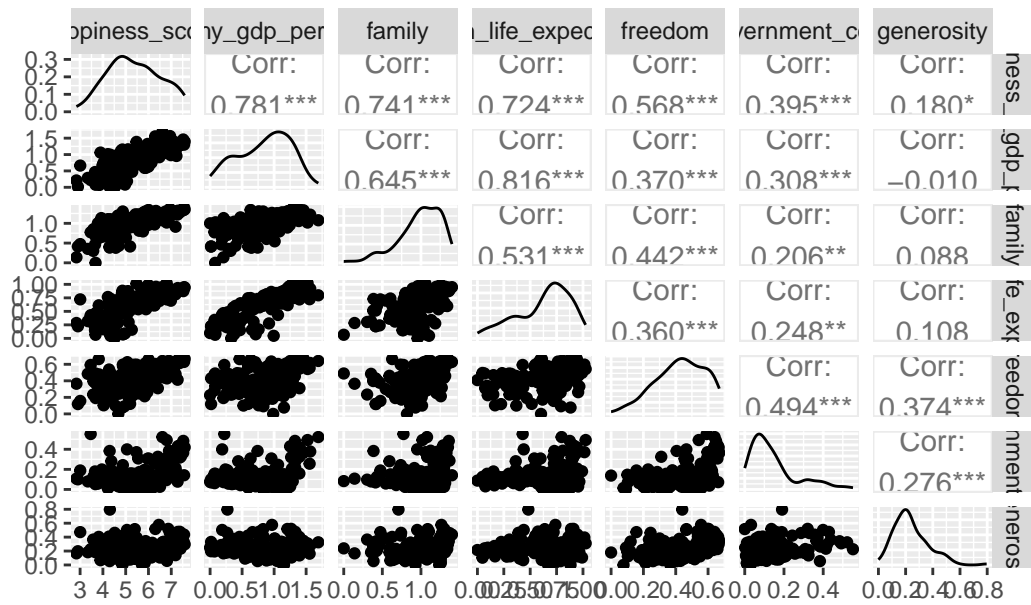
```
method from  
+.gg    ggplot2
```

Attaching package: 'GGally'

The following object is masked \_by\_ '.GlobalEnv':

happy

## Pair-wise Relationships Among Key Metrics



## Appendix

```
{r codeAppend, ref.label=knitr::all_labels(), echo=TRUE, eval=FALSE}
```