Researching Global Happiness

David Booker-Earley

Presentation Outline

- Intro
  - Data overview & Goal
- Exploration
  - Hypotheses & Analysis
- Results
- Next
  - Future research
- Appendix
  - Assumptions & Equations
  - Supplemental data

The World Happiness Report

- Evaluates happiness for 156 countries
- Helps assess nations' progress
- 2019 data has six factors
- Located on Kaggle

“Factor Values” from 6 Factors of Happiness

- GDP per capita (Economic Productivity)
- Social support (from Family, Friends, etc.)
- Healthy life expectancy
- Freedom to make life choices
- Generosity
- Perceptions of corruption

Wait,

... what?

Hi!

... I'm the ~~(pseudo)~~ on-screen assistant.

... How can I help?

Uh,

... factor values?

Basically,

... they tell us how "valuable" each factor is

... but for each country.

Oh, okay, sounds cool!

... But which factors make countries the happiest?

Good Question!

... Let's find out!

Goal

➢ Determine how given factors contribute to happiness.

How?

1 → Hypothesize

- *.. Make an educated guess!*

2 → Investigate

- *.. Test that educated guess!*

3 → Analyze Results

- *Explain stuff!*

Forming 3 Hypotheses

Investigative Questions

1. *Does a country's GDP per capita make people happier than having a healthy life expectancy?*
2. *Is a country's perceived corruption related to the overall happiness of its citizens?*
3. *Does having social support make people happier than having the freedom to make life choices?*

Which factor will have the higher average value?

H.1

Expected

economic productivity

vs

healthy life expectancy

H.3

Expected

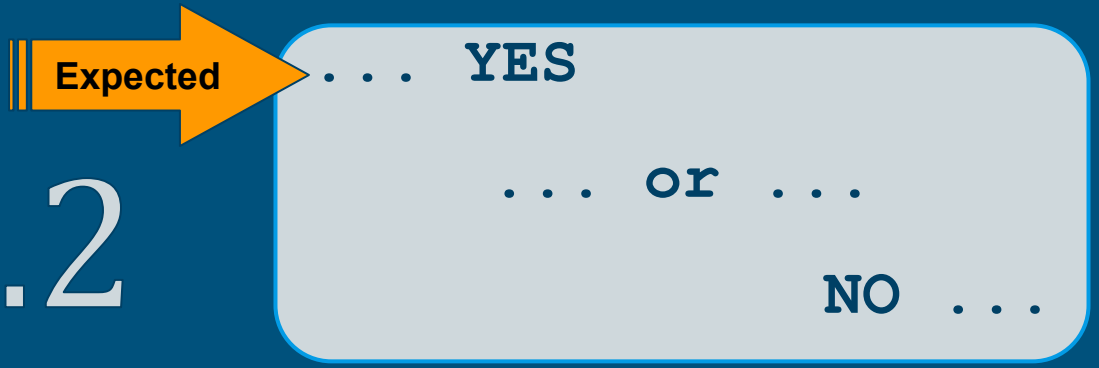
social support

vs

freedom to make life choices

1st & 3rd Questions

Is perceived corruption related to overall happiness?



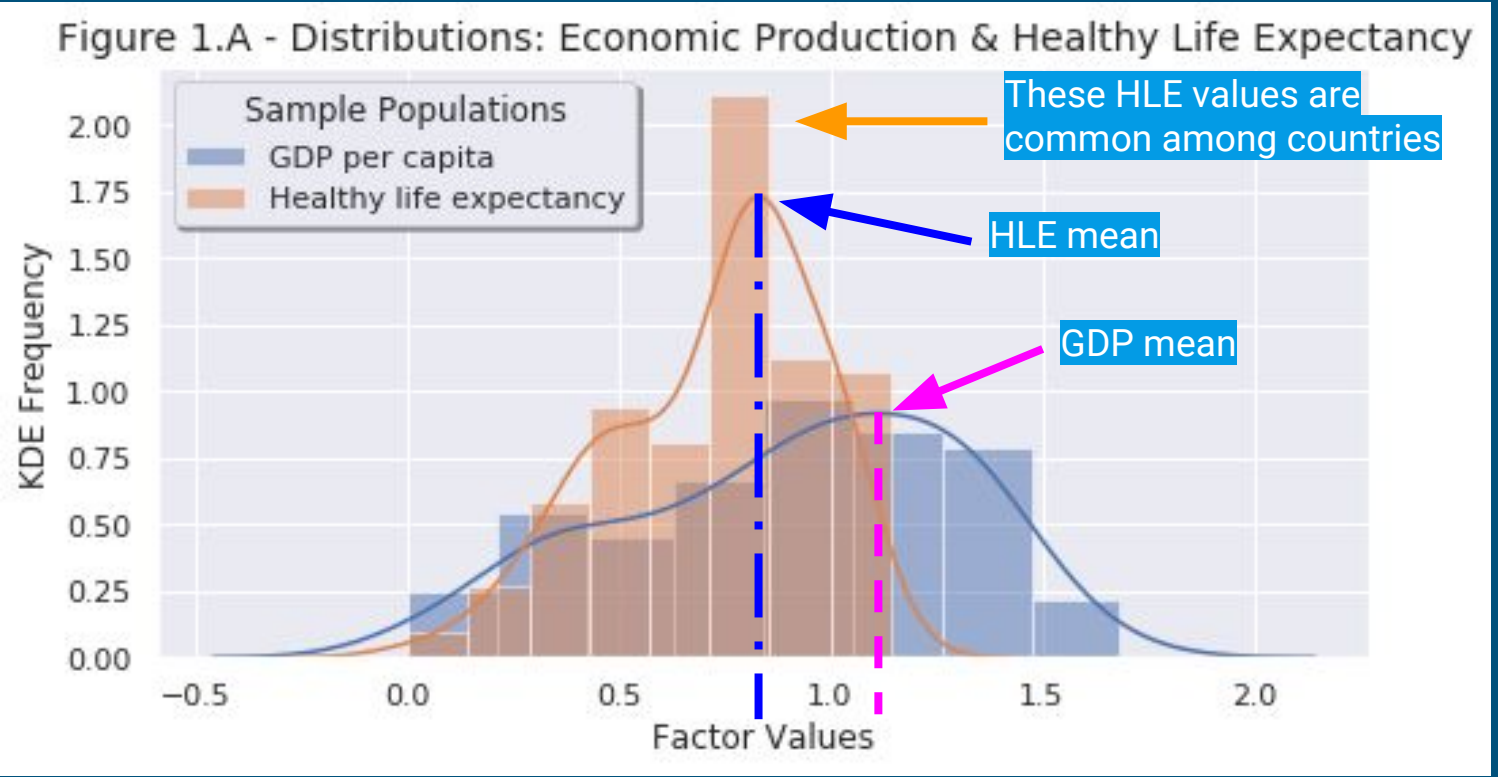
Cool, we've guessed stuff

... now how do we do the other stuff?

With a little help from some old friends,

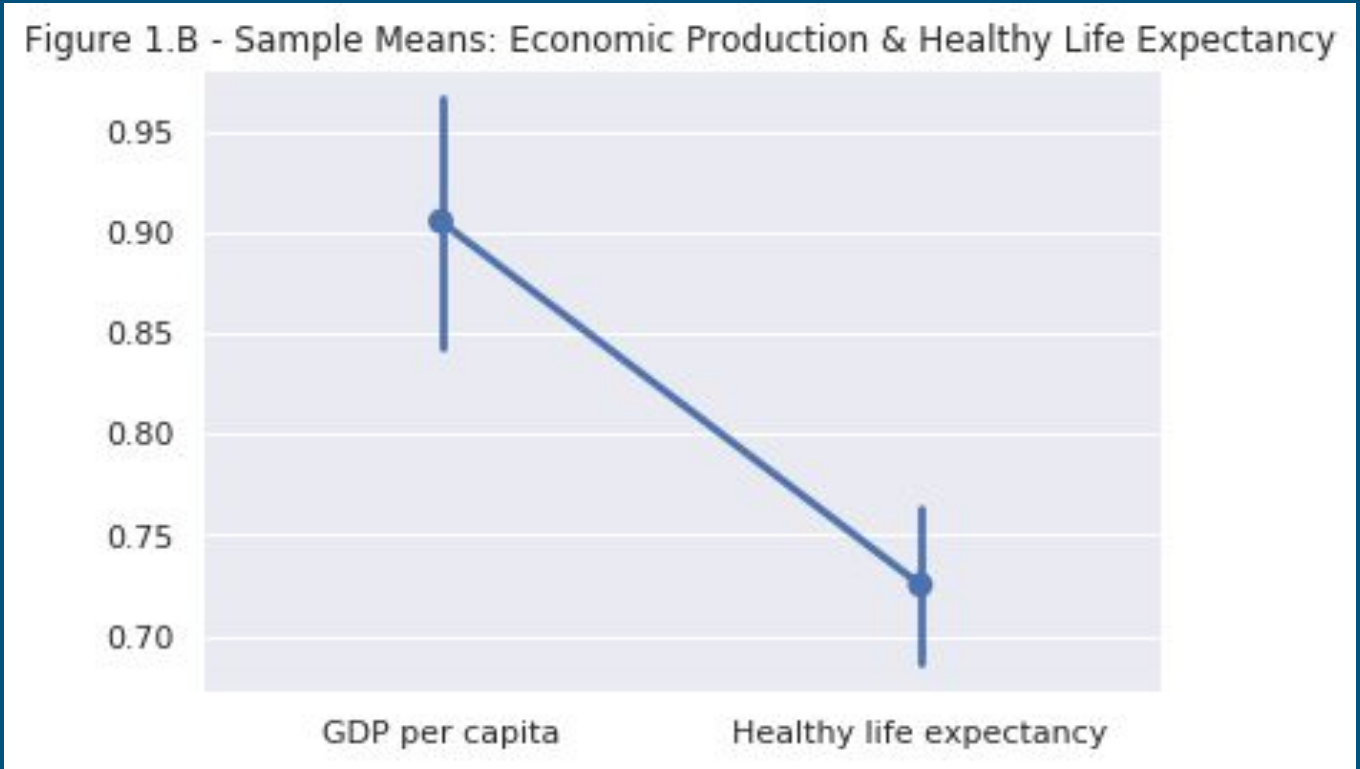
... Math & Pictures!

Exploring 3 Hypotheses



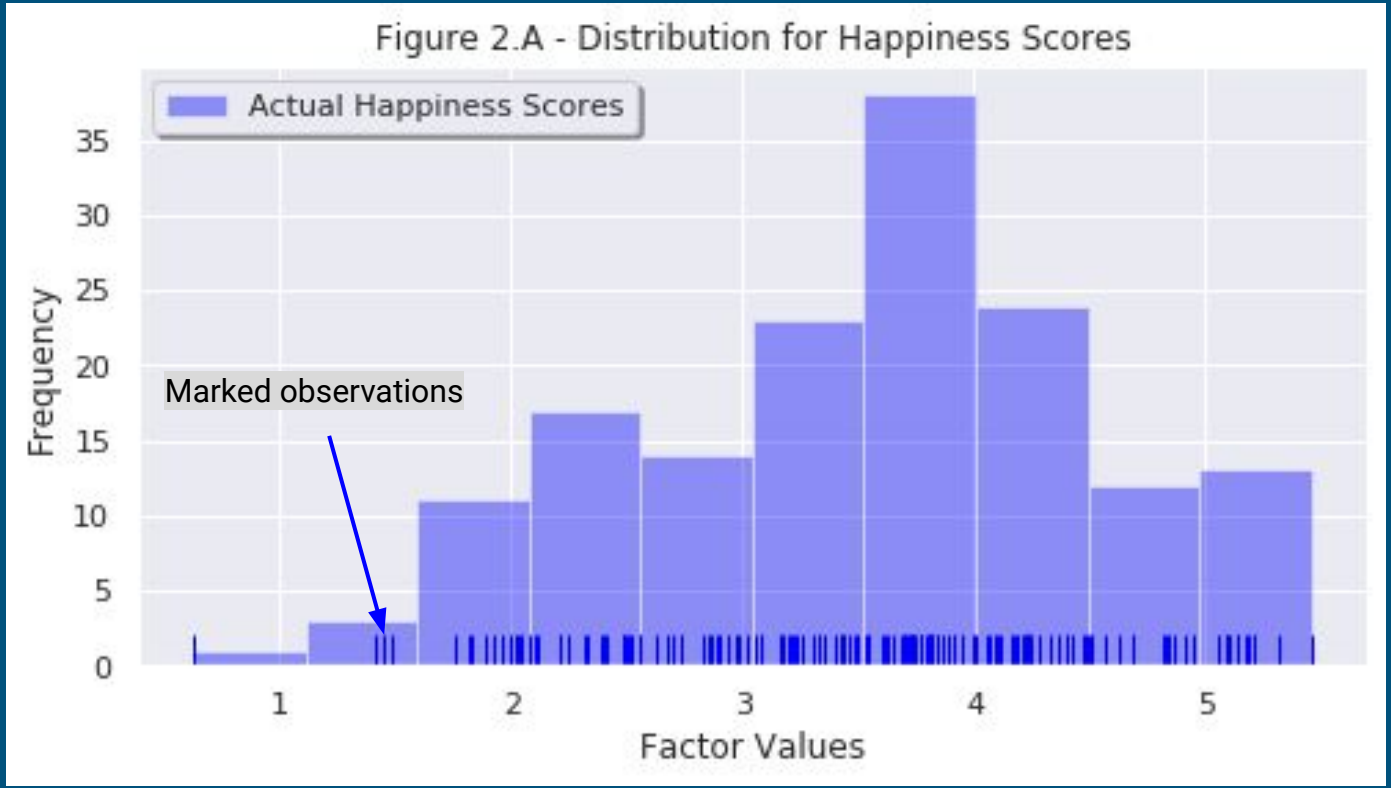
H.1

Most values exist between 0.25 and 1.5



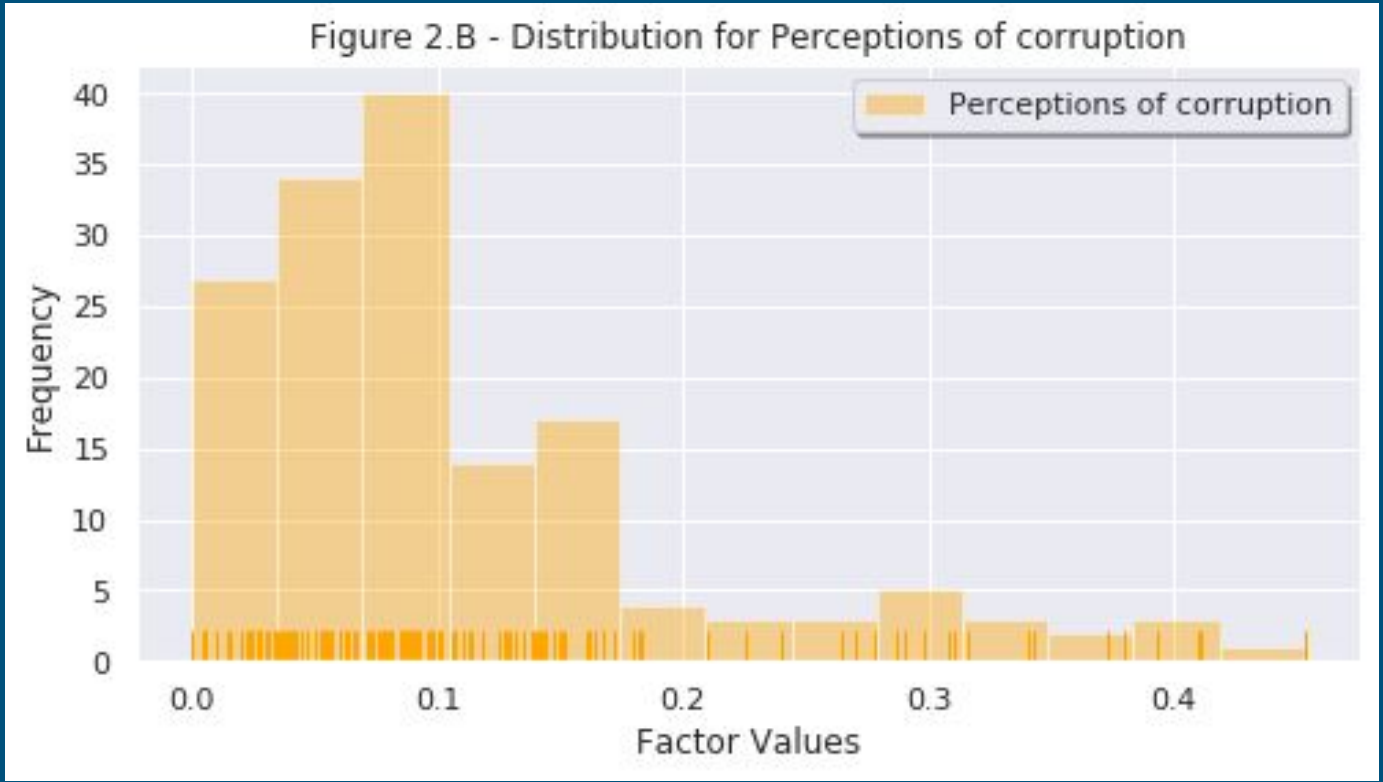
H.1

GDP per Capita mean is greater than HLE mean



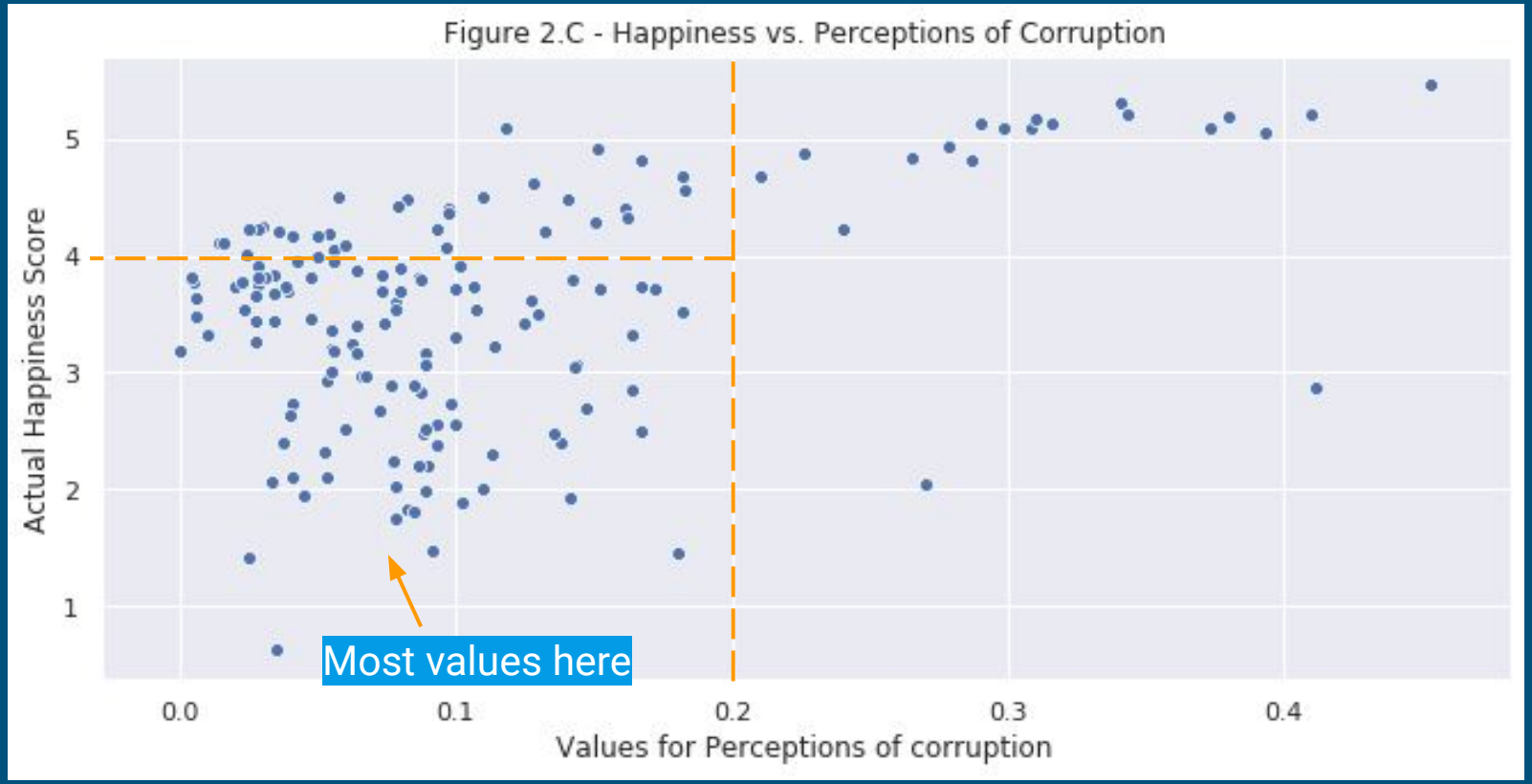
H.2

Most values exist below 4.5



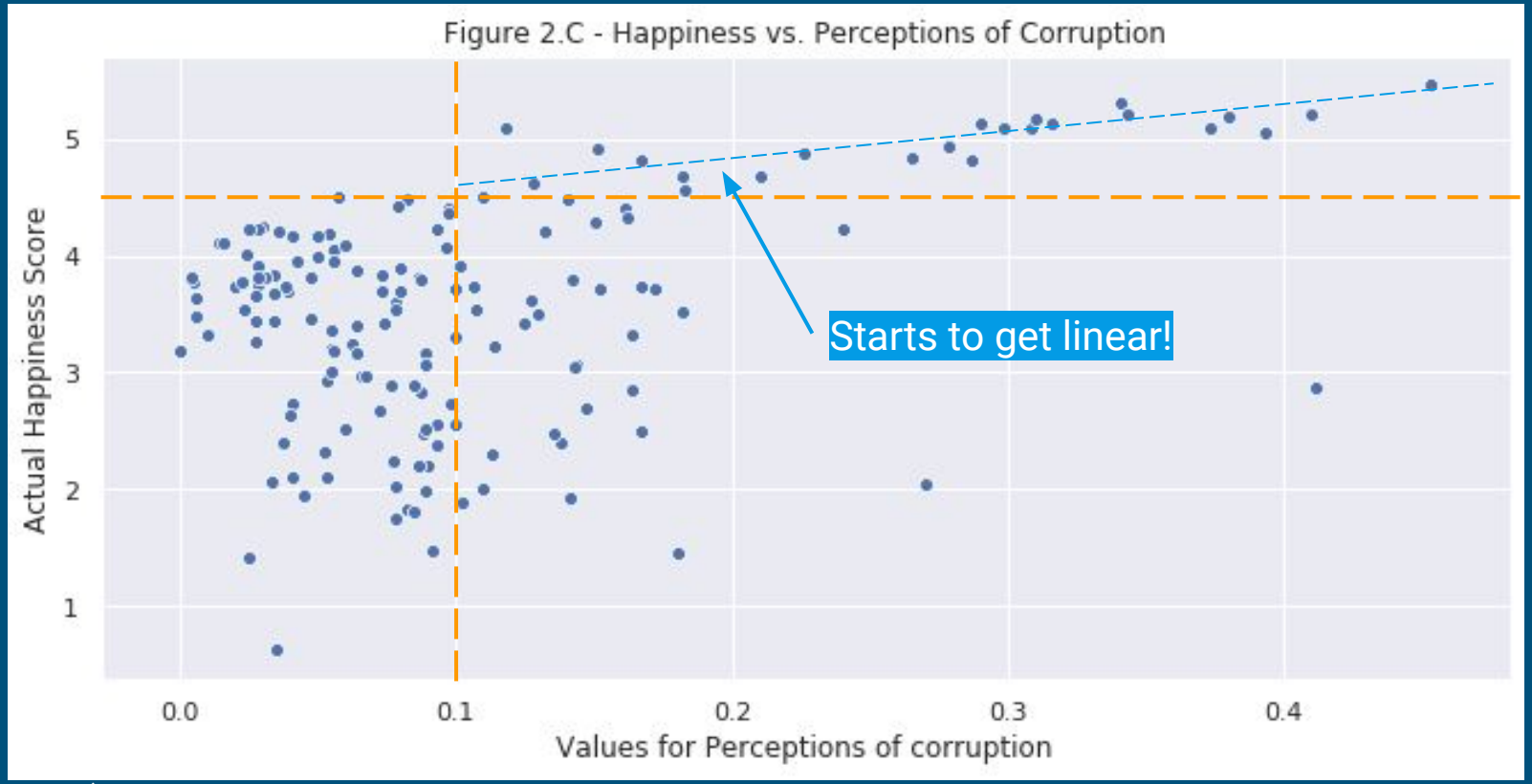
H.2

Most countries had low values for POC



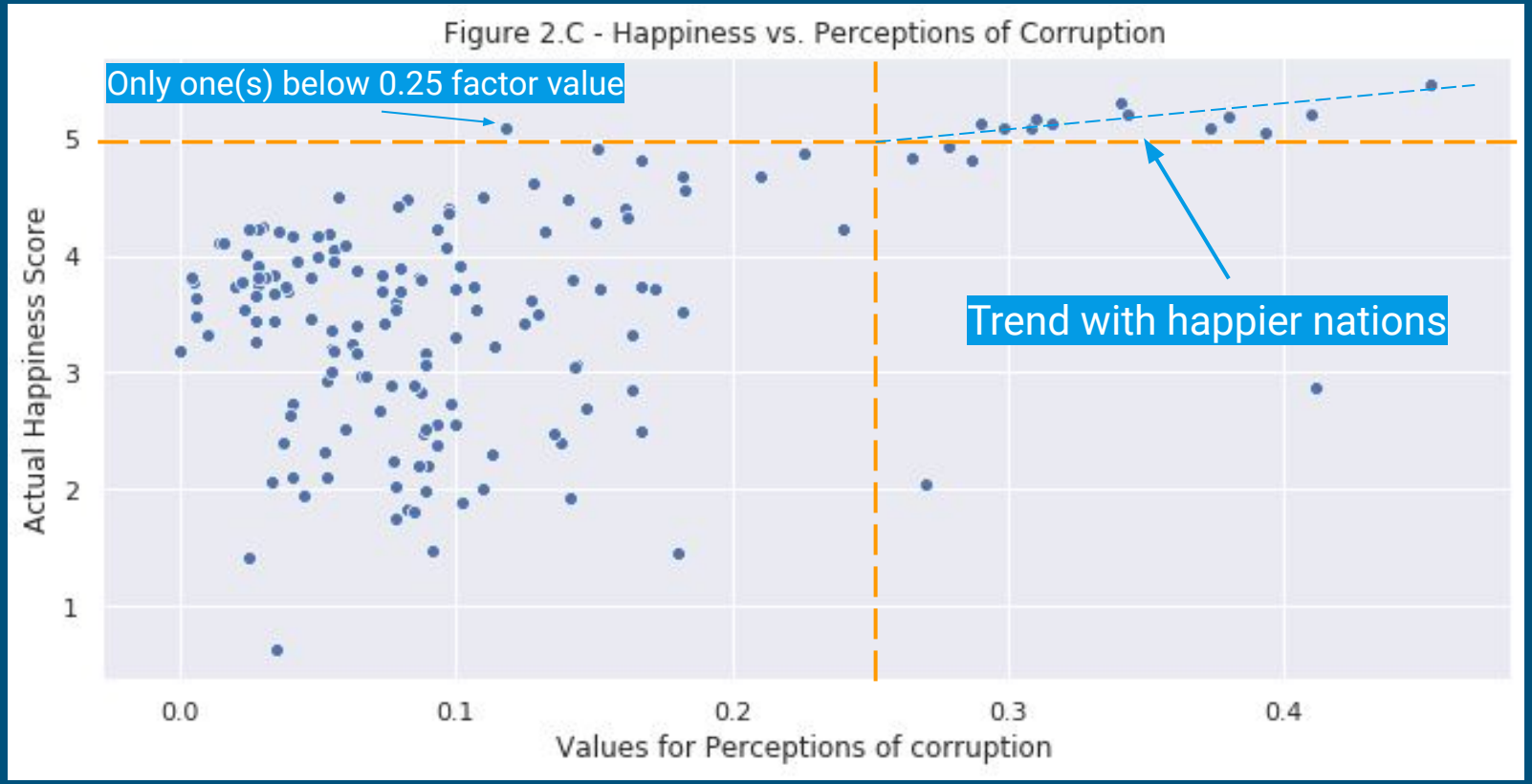
- 1 → Most perceived corruption values that are less than 0.2 were for countries with happiness scores below 4.0

H.2



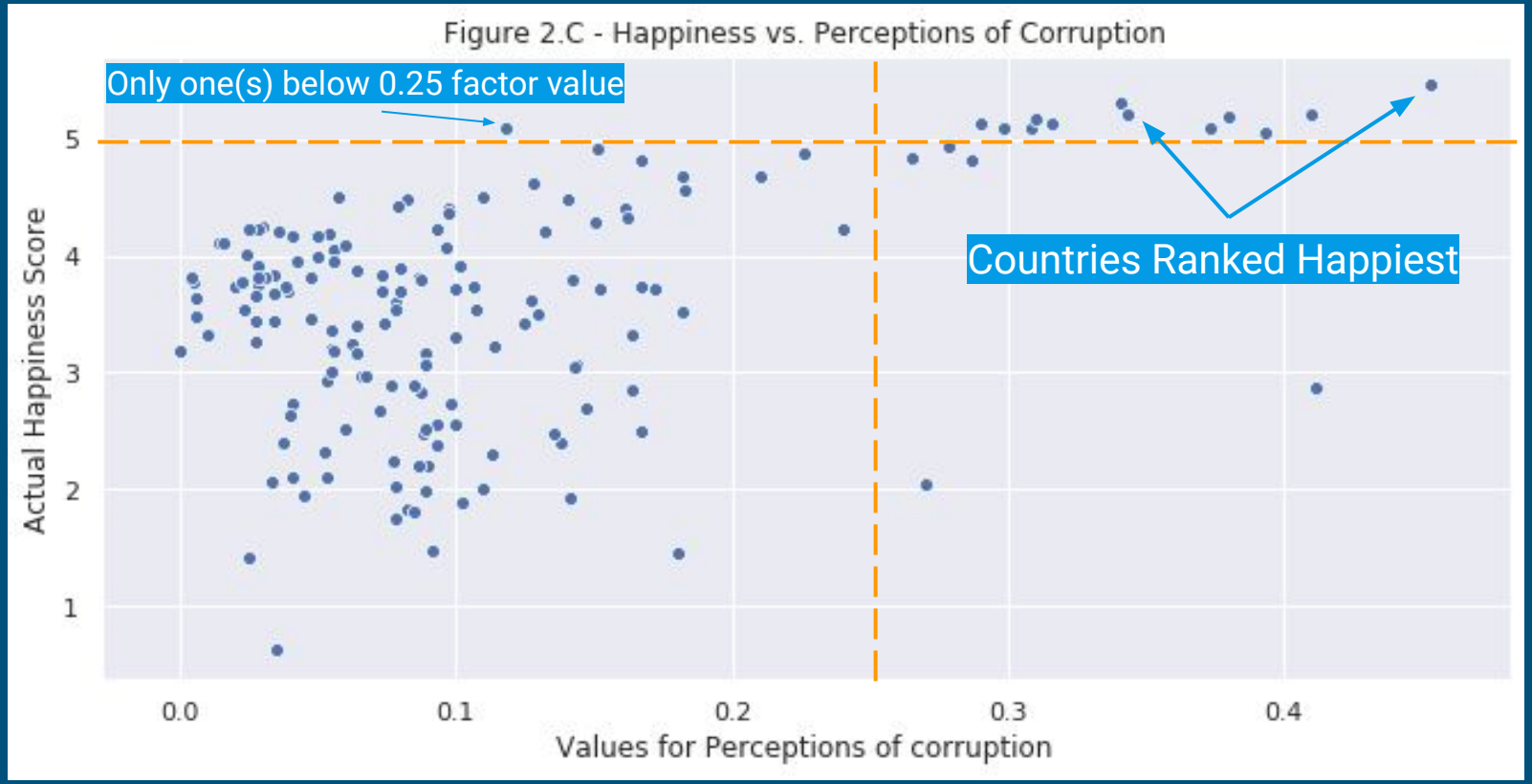
- 2 → Considerable linearity exists after perceived corruption values just after 0.1 for happiness scores above 4.5

H.2



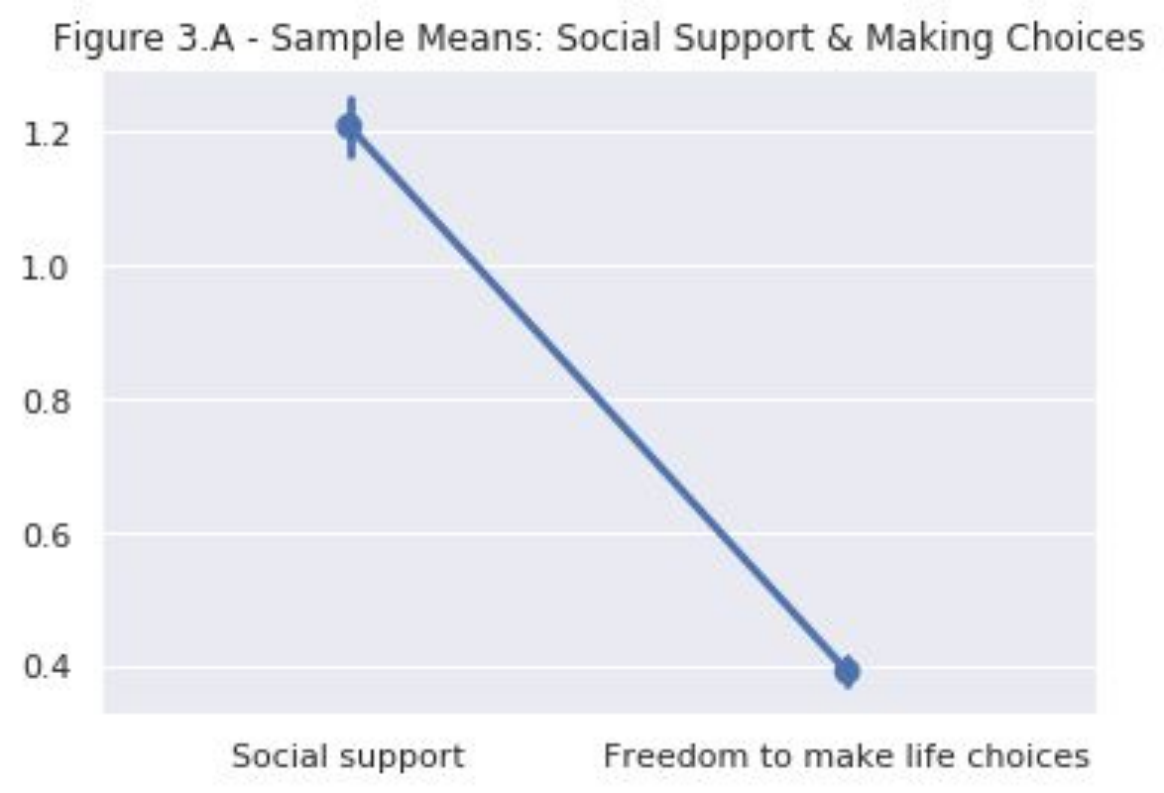
- 3 → Almost all happiness scores larger than 5.0 had perceived corruption values greater than 0.25

H.2



- Happiest nations consistently saw more value in this factor than all lower-ranking nations did.

H.2



H.3

- Is there significant difference between the two averages? → Yes
- Is average for SS higher than avg for FMLC? → Yes

Quantitatively, by how much do they differ?

With 95% Confidence,

Difference in Factor Values exists between 0.76 and 0.87

H.3

Visualize Results with “%”

- Social support → 34% (of contribution to happiness)
- Freedom to make life choices → 11%

23 % Difference in Factor Values!

H.3

Uh, cooooool

... Sorry, I got lost :(

... How does all of this answer those 3 "detective questions" from earlier?

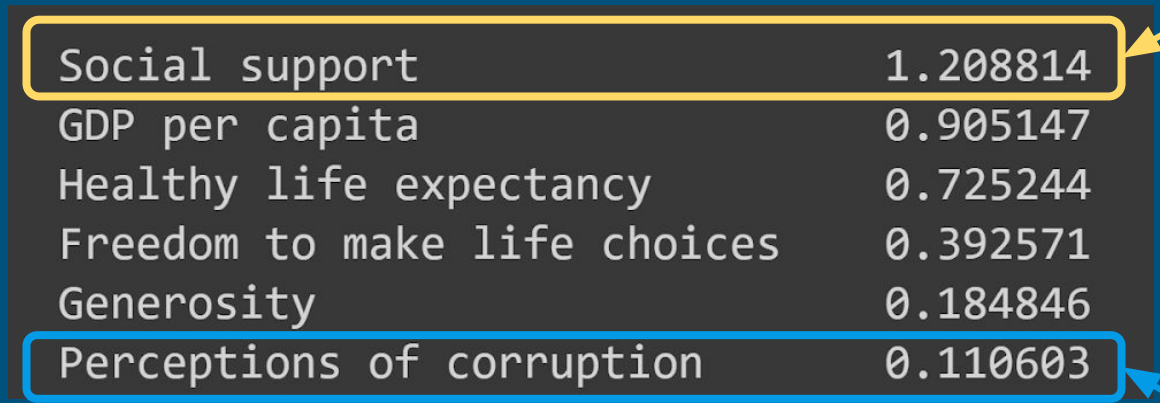
No worries!

... We have the ingredients,

... so let's put everything together to see!

Results in Context

Factor Averages

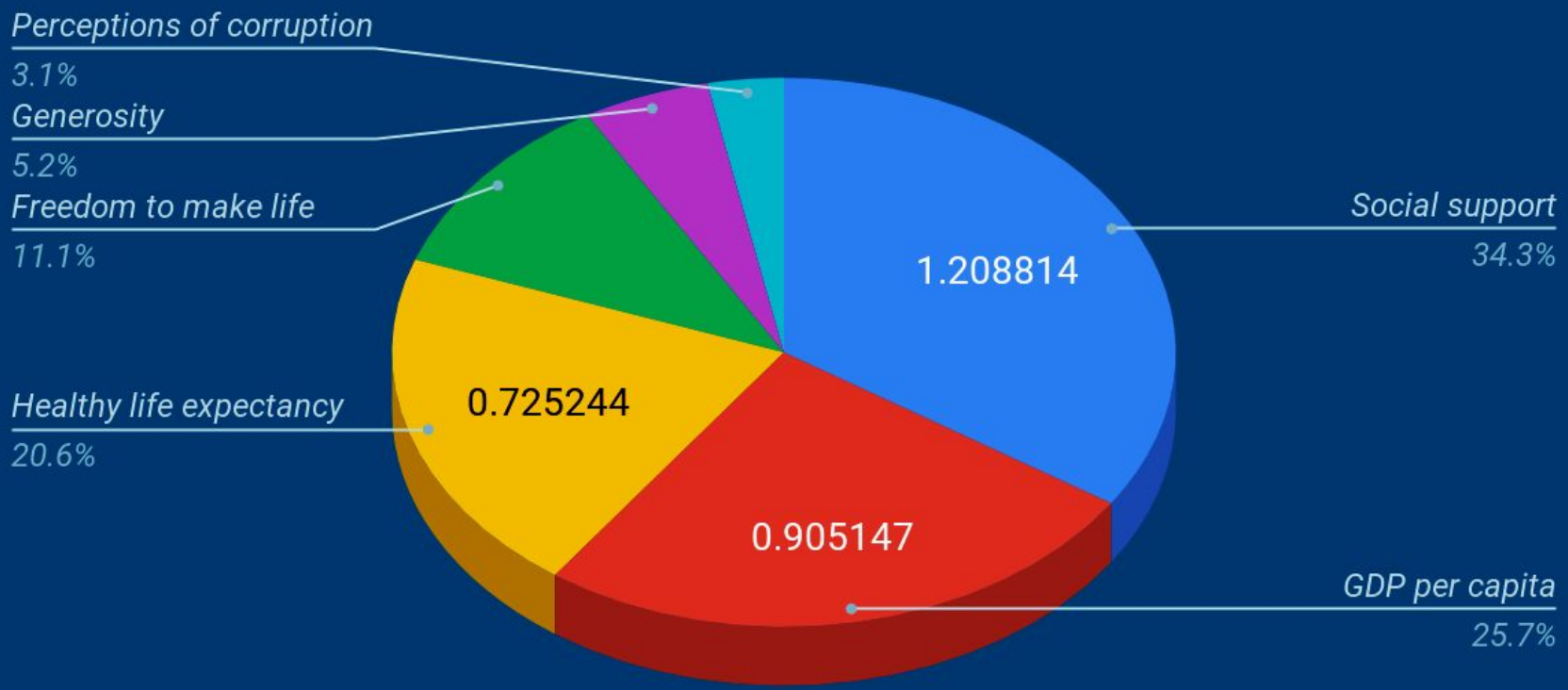


Social support	1.208814
GDP per capita	0.905147
Healthy life expectancy	0.725244
Freedom to make life choices	0.392571
Generosity	0.184846
Perceptions of corruption	0.110603

Highest

Lowest

Average Value per Happiness Factor



Top & Bottom Contributors

- Social support → 34%
- Perceptions of corruption → 3%

31 % Difference in Factor Values!

What about those Hypotheses?

H.1 → **Accepted** → ***economic productivity*** .. contributed more than .. healthy life expectancy

H.3 → **Accepted** → social support .. contributed more than .. freedom to make life choices

What happened with the 2nd Hypothesis?

H.2

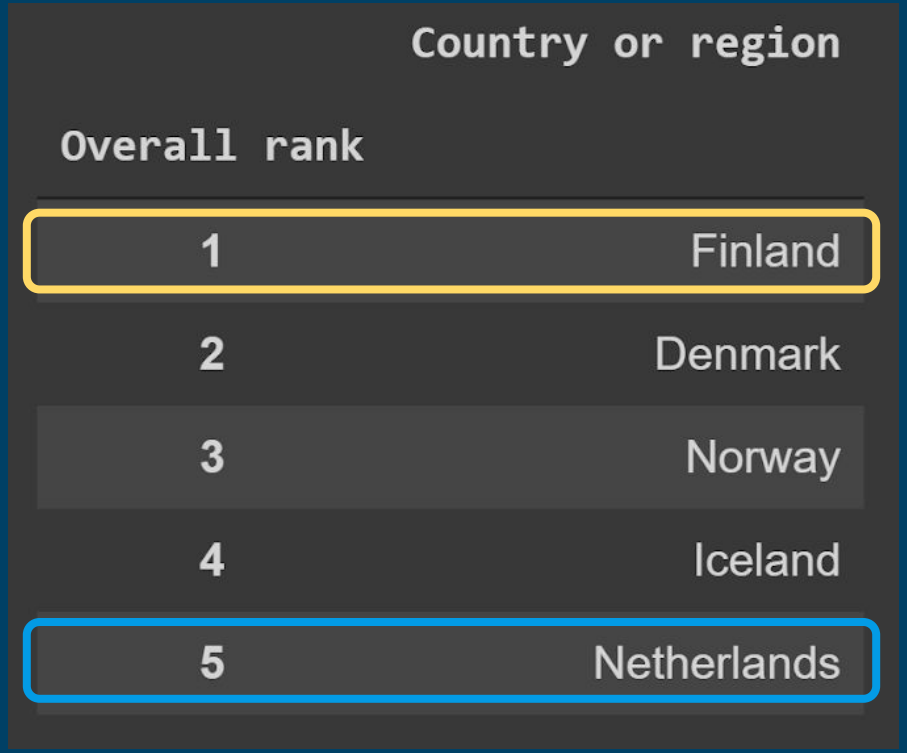
The data says:

- Overall, “perceived corruption” did not contribute much to happiness.
- But, happier nations valued it more than low-ranking nations did.

So, who is the happiest?

Top 5 Ranks

Countries with the highest happiness scores



Overall rank	Country or region
1	Finland
2	Denmark
3	Norway
4	Iceland
5	Netherlands

Bottom 5 Ranks

Countries with the lowest happiness scores

Overall rank	Country or region
152	Rwanda
153	Tanzania
154	Afghanistan
155	Central African Republic
156	South Sudan

Cool, I think I get it now

... but aren't there other factors that could go into all of this?

Another Great Question!

... Shortest Answer: Yes!

Longer answer:

... health conditions
... societal constructs
... tech accessibility

... bias in the data collection process

... the list can go on!

Wow, I see now, thanks!

... that was pretty fun!

... So, like "irl", who can use these results?

Economists, Governments, Policy makers, Psychologists, Statisticians,

... and, of course,
... Data Scientists! :)

That's pretty cool!

... what could they do with all of this stuff?

Shortest Answer:

1 ... Compare happiness trends of top-ranking nations with their own.

2 ... Identify strengths & weakness for progress!

Next Steps:

Next Steps for Further Research

1. Test the same hypotheses for data from other years (2015, 2016, etc.).
2. Calculate and examine correlations between various factors.

Next Steps for Further Research

1. Test the same hypotheses for data from other years (2015, 2016, etc.).
2. Calculate and examine correlations between various factors.

> For the highest-ranking and lowest-ranking countries, respectively:
>
> 3. Determine how factor values changed over time.
> 4. Outline the trend or oscillations of a country's overall rank.

Next Steps for Further Research

1. Test the same hypotheses for data from other years (2015, 2016, etc.).
2. Calculate and examine correlations between various factors.

> For the highest-ranking and lowest-ranking countries, respectively:
>
> 3. Determine how factor values changed over time.
> 4. Outline the trend or oscillations of a country's overall rank.
> 5. Research various aspects of how daily-life differs (perhaps using the variables and secondary metrics mentioned earlier).
> 6. Based on those research results, discuss how effective each country is at improving its overall happiness; perhaps relate it to interesting patterns observed (like with healthcare, technological developments, work-life balance, etc.).

Thank you for your time!

Any questions?

Appendix

```
Sample population mean for actual happiness scores:
3.527224

Sample population mean per factor, sorted from greatest to least:

Social support                    1.208814
GDP per capita                    0.905147
Healthy life expectancy           0.725244
Freedom to make life choices      0.392571
Generosity                        0.184846
Perceptions of corruption         0.110603
```

Sample Population Mean per Variable

Overall rank	Country or region	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Actual Happiness Score
1	Finland	7.769	1.340	1.587	0.986	0.596	0.153	0.393	5.055
2	Denmark	7.600	1.383	1.573	0.996	0.592	0.252	0.410	5.206
3	Norway	7.554	1.488	1.582	1.028	0.603	0.271	0.341	5.313
4	Iceland	7.494	1.380	1.624	1.026	0.591	0.354	0.118	5.093
5	Netherlands	7.488	1.396	1.522	0.999	0.557	0.322	0.298	5.094

Data for Top 5 ranks

Overall rank	Country or region	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Actual Happiness Score
152	Rwanda	3.334	0.359	0.711	0.614	0.555	0.217	0.411	2.867
153	Tanzania	3.231	0.476	0.885	0.499	0.417	0.276	0.147	2.700
154	Afghanistan	3.203	0.350	0.517	0.361	0.000	0.158	0.025	1.411
155	Central African Republic	3.083	0.026	0.000	0.105	0.225	0.235	0.035	0.626
156	South Sudan	2.853	0.306	0.575	0.295	0.010	0.202	0.091	1.479

Data for Bottom 5 ranks

Assumptions for using the `t-test`

1. The test variable is continuous.
2. There is no relationship between items; that is, the measurement of one observation does not affect the measurement of another.
3. Samples are drawn at random for each population.
4. The test variable's samples and populations are approximately normally distributed.
5. Reasonably large enough samples were collected to be representative of the population.
6. Variances are approximately equal in both the sample and population.

Result thresholds for the `t-test`

- At the 95% Confidence Interval (two-tail)
  - The critical value for the `test statistic` is ∓ 1.96
  - The critical value for the `p-value` is 0.05 or 5%

Notes for formal `t-test` | Part 1

Formula used to perform the `t-test`

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

- $\bar{x}_1$ is the `mean` of the first sample set
- $\bar{x}_2$ is the `mean` of the second sample set
- s_1 is the `standard deviation` of the first sample set
- s_2 is the `standard deviation` of the second sample set
- n_1 is the `sample size` of the first sample set
- n_2 is the `sample size` of the second sample set

Notes for formal `t-test` | Part 2

Formula used to calculate the 95% Confidence Interval

$$\bar{x}_1 - \bar{x}_2 \mp 1.96 * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where:

- $\bar{x}_1$ is the `mean` of the first sample set
- $\bar{x}_2$ is the `mean` of the second sample set
- s_1 is the `standard deviation` of the first sample set
- s_2 is the `standard deviation` of the second sample set
- n_1 is the `sample size` of the first sample set
- n_2 is the `sample size` of the second sample set

Equation for `95% Confidence Interval (two-tail)`

- Statistical Insight → `t-test` says:
  - There is a significant statistical difference between the two population means.

```
Ttest_indResult(statistic=30.73210125606325, pvalue=4.142244084525088e-96)
```

p-value << 0.001%

social support

.. contributes more than ..

freedom to make life choices

- Substantive Insight → 95% Confidence Interval

```
The difference in means at the 95% Confidence Interval (two-tail) is between 0.76 and 0.87.
```

Results from Formal Statistical Test

Other Variables

- Evaluating `overall happiness` by tracking other variables, such as the following, could help outline patterns between happier, healthier, and more successful countries:
  - Group 1 Variables: Data Overviews
    - *Employment rates*
    - *Famine rates*
    - *Homelessness rates*
    - *Medical outbreaks per unit of time*
    - *Poverty rates*
    - *Technological or scientific developments per unit of time*
    - *Treaties or disputes with other nations*
  - Group 2 Variables: Data from Individuals
    - *Satisfaction with occupation or financial status*
    - *Accessibility of "opportunities to make life choices for improvement"*
    - *Other health factors (like various types of depression or trauma)*

Other Variables (factors that affect happiness)

Further research is needed to see how and why countries may fall or rise in the ranks over several years. The combined research results would enable interested parties to make research-driven decisions and track the progress of implementation.

> Knowing how high-ranking countries maintain happiness while achieving progress would help interested parties do, for example, the following:
>
> - Identify new ways to improve social morale.
> - Distinguish alternatives for healthier life-balances (exercise, diet, work-play, etc.).
>   - Simultaneously, data from these actions would provide more insight to how `happier citizens` can improve `overall productivity` between various businesses locally and internationally.
> - Explore innovative solutions to mitigate persisting challenges (emissions, energy sources, extreme weather and wildfires, healthcare, pollution, poverty, recycling, distribution of wealth, etc.).

Future Research Insight

Recommended Updates for Data Collection

According to Kaggle, `happiness scores and rankings use data from the Gallup World Poll.` However, the survey's implementation is unclear. For future data collection processes, tracking some or all of the following items could help avoid adverse effects in data analysis and interpretation:

- Diversity
  - *Who was surveyed; were they selected at random (activist, civil workers, doctors, educators, engineers, government officials, musicians, parents, politicians, researchers, students, etc.)?*
- Duration
  - *At which time of the year were surveys given and submitted / completed, respectively?*
- Quality
  - *How were the surveys distributed (via email, in person, phone, etc.)?*
- Quantity
  - *How many people were surveyed to represent the desired population?*

> These questions outline how the sample population may differ from the desired (actual) population, along with secondary metrics that could provide insight into the observed differences. Without knowing the answers to these questions, it's unclear how one could test for either *observer bias* and *sampling / selection bias*.

Data Collection Recommendations

Next Steps for Further Research

1. Test the same hypotheses for data from other years (2015, 2016, etc.).
2. Calculate and examine correlations between various factors.

> For the highest-ranking and lowest-ranking countries, respectively:
>
> 3. Determine how factor values changed over time.
> 4. Outline the trend or oscillations of a country's overall rank.
> 5. Research various aspects of how daily-life differs (perhaps using the variables and secondary metrics mentioned earlier).
> 6. Based on those research results, discuss how effective each country is at improving its overall happiness; perhaps relate it to interesting patterns observed (like with healthcare, technological developments, work-life balance, etc.).

Next Steps, Further Research

Sources

1. Kaggle Datasets
    a. https://www.kaggle.com/unsdsn/world-happiness#2019.csv
2. Picture of globe
    a. Image by OpenClipart-Vectors from Pixabay
3. Pictures of emoji
    a. Emoji 1 → Image by Ralf Uy from Pixabay
    b. Emoji 2 → Image by Ralf Uy from Pixabay