A Statistical Exploration of Happiness

UNIVERSITY of WASHINGTON

Team Members

- Isaac Kim
- Urmika Kasi
- Andy Tsai
- Abhishek Saini
- Kevin Torrico

Learning Objectives

- Discovering the important factors in determining happiness
- Analyzing impacts and correlations Covid-19 had on happiness levels around the world
- Understanding and comparing happiness levels between different geographical regions

Dataset Selection

- https://worldhappiness.report/ed/2021/
- The Gallup World Poll data was used to calculate the happiness scores and rankings
- Six factors – economic production, social support, life expectancy, freedom, lack of corruption, and generosity
- Six geographical regions

How does each feature affect happiness?

- Method Used: Linear Regression test for each feature with ladder score as the response variable and each feature as their own predictor variable.

- Null Hypothesis: H0 : β1 = 0 GDP per Capita
  β2 = 0 Social Support
  β3 = 0 Healthy Life Expectancy
  β4 = 0 Freedom to make life choices
  β5 = 0 Generosity
  β6 = 0 Perception of corruption

Linear Regression Results

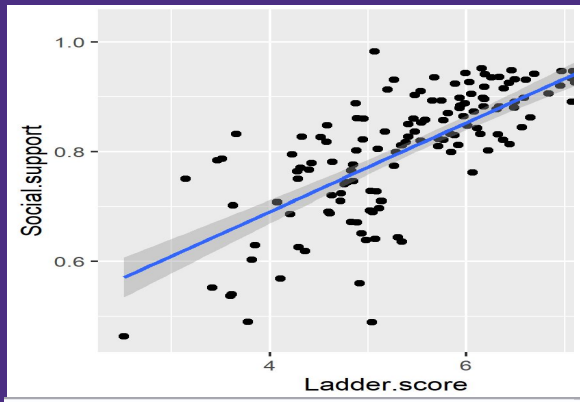
- There is sufficient evidence to reject the null hypothesis for the following:

p-value: < 2.2e-16	β1 != 0 GDP per Capita
p-value: < 2.2e-16	β2 != 0 Social Support
p-value: < 2.2e-16	β3 != 0 Healthy Life Expectancy
p-value: < 2.2e-16	β4 != 0 Freedom to make life choices
p-value: 8.881e-08	β6 != 0 Perception of corruption

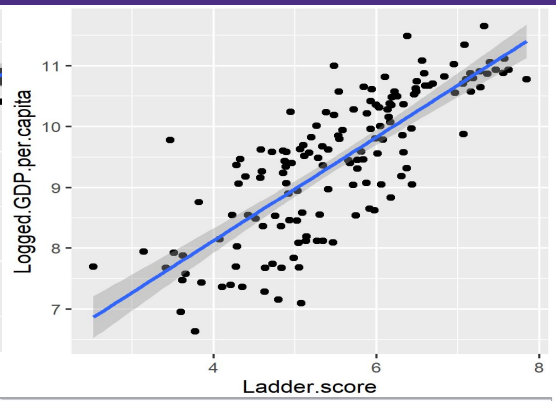
- We fail to reject the null hypothesis for:

p-value: 0.8294	β5 = 0 Generosity

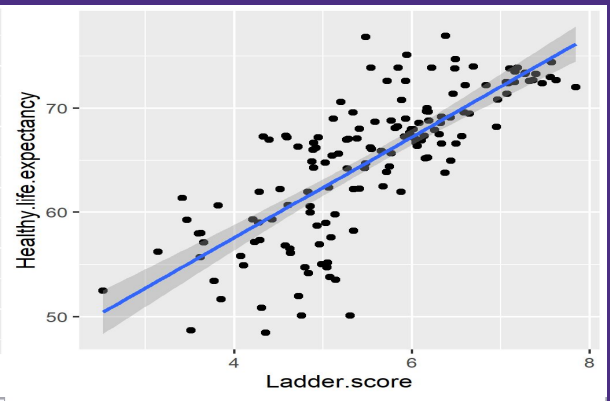
Graphical Results



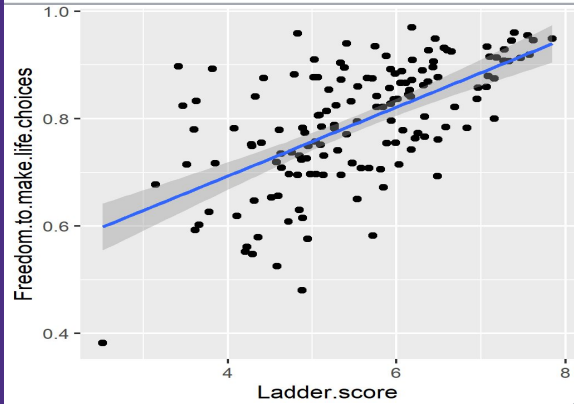
$$LadderScore = -0.2315 + 7.075 * SocialSupport$$



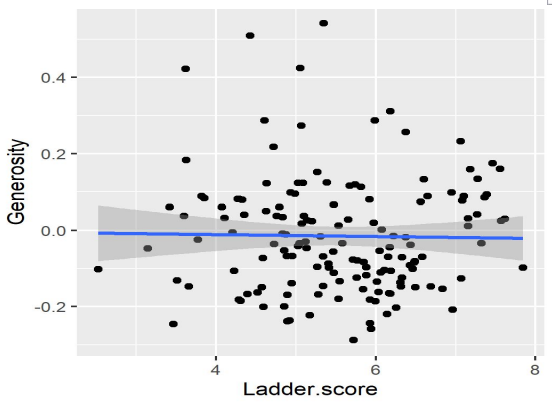
$$LadderScore = -1.3719 + 0.732 * GDPperCapita$$



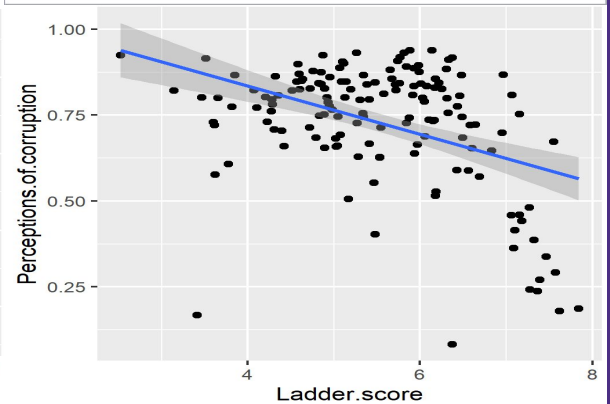
$$LadderScore = -2.395 + 0.121 * HealthyLifeExpectancy$$



$$LadderScore = 0.974 + 5.759 * FreedomLifeChoices$$



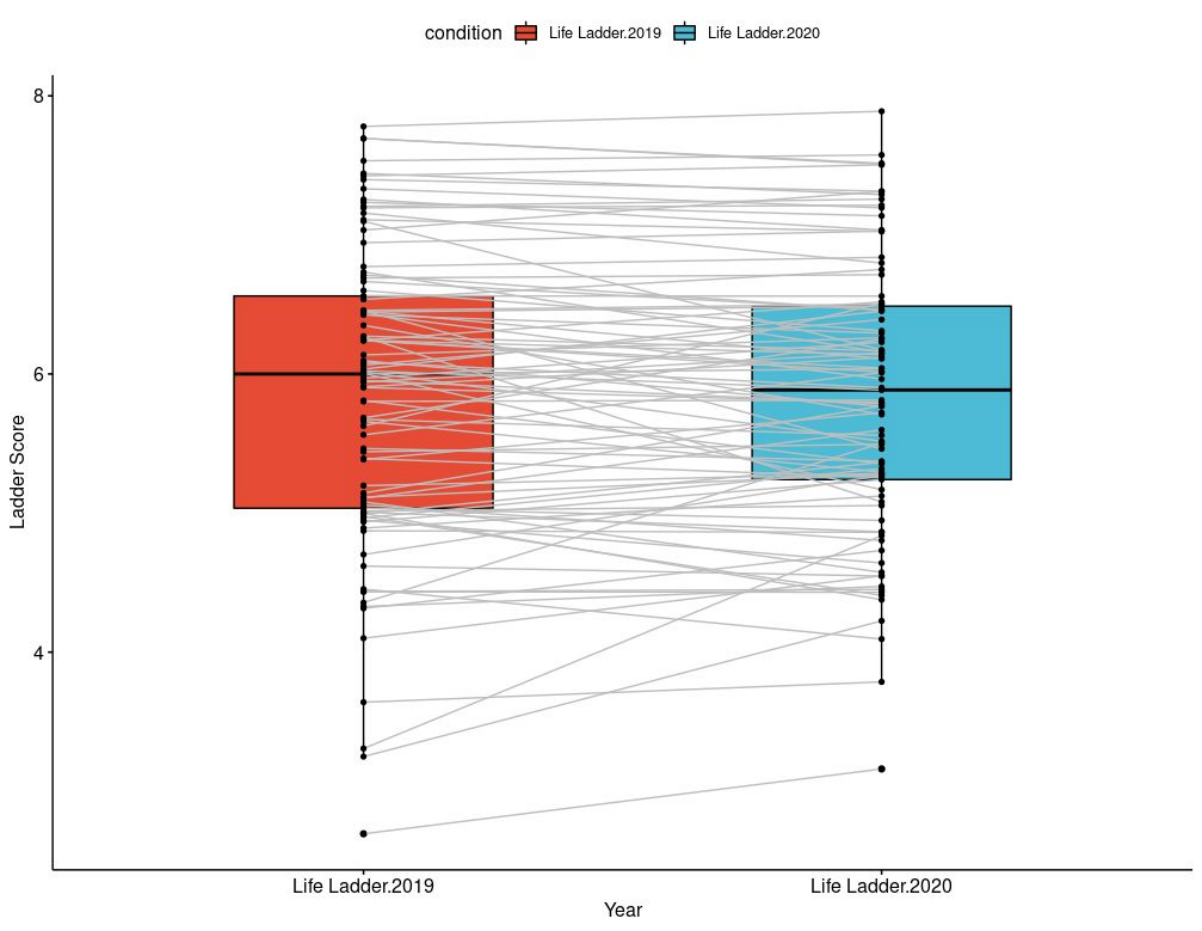
$$LadderScore = 5.53 - 0.126 * Generosity$$



$$LadderScore = 7.368 - 2.523 * PerceptionsOfCorruption$$

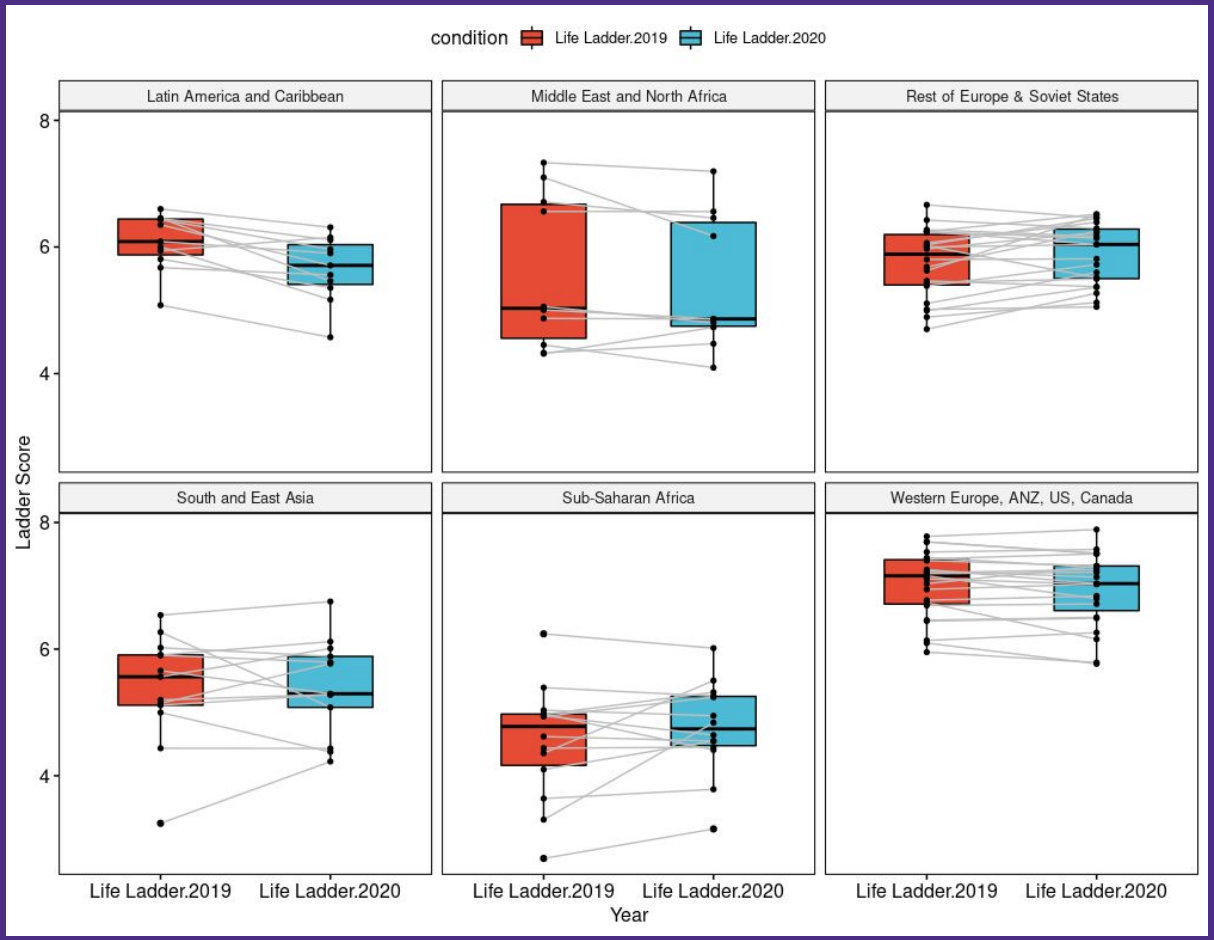
Impact of COVID-19 on happiness scores?

Data Viz. Change in Ladder Score from 2019 to 2020 - all countries



Impact of COVID-19 on happiness scores?

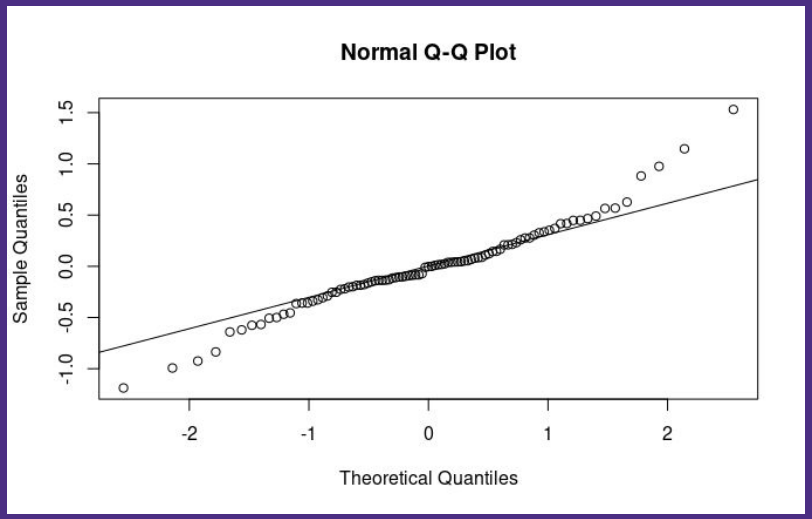
Data Viz. Change in Ladder Score from 2019 to 2020 - by region



Impact of COVID-19 on happiness scores?
Hypothesis Testing

- Paired t-test
- Null Hypothesis
  - For all the countries, ladder scores are the same before and after the pandemic. $H_0 : \mu_{2019} = \mu_{2020}$ versus $H_1 : \mu_{2019} \neq \mu_{2020}$
  - For every region, ladder scores are the same before and after the pandemic. $H_0 : \mu_{2019} = \mu_{2020}$ versus $H_1 : \mu_{2019} \neq \mu_{2020}$

Impact of COVID-19 on happiness scores?
Validation of Assumptions for paired t-test



> Normality assumption doesn't hold and outliers are present

Impact of COVID-19 on happiness scores?
Results

> We find a statistically significant drop in happiness scores in 'Latin America and Caribbean' region

region	n	mean_2019	mean_2020	change	p_val_t_test	p_val_wilc
Rest of Europe & Soviet States	22	5.74	5.90	0.16	0.0303	0.0425
Latin America and Caribbean	11	6.08	5.66	-0.42	0.0020	0.0049
Western Europe, ANZ, US, Canada	23	7.02	6.95	-0.07	0.1222	0.1792
Middle East and North Africa	10	5.57	5.42	-0.15	0.2051	0.1934
South and East Asia	13	5.38	5.41	0.03	0.8802	0.6848
Sub-Saharan Africa	14	4.55	4.76	0.21	0.1783	0.2958
Overall	93	5.85	5.84	-0.01	0.8536	0.5926

Are all regions of the world equally happy?

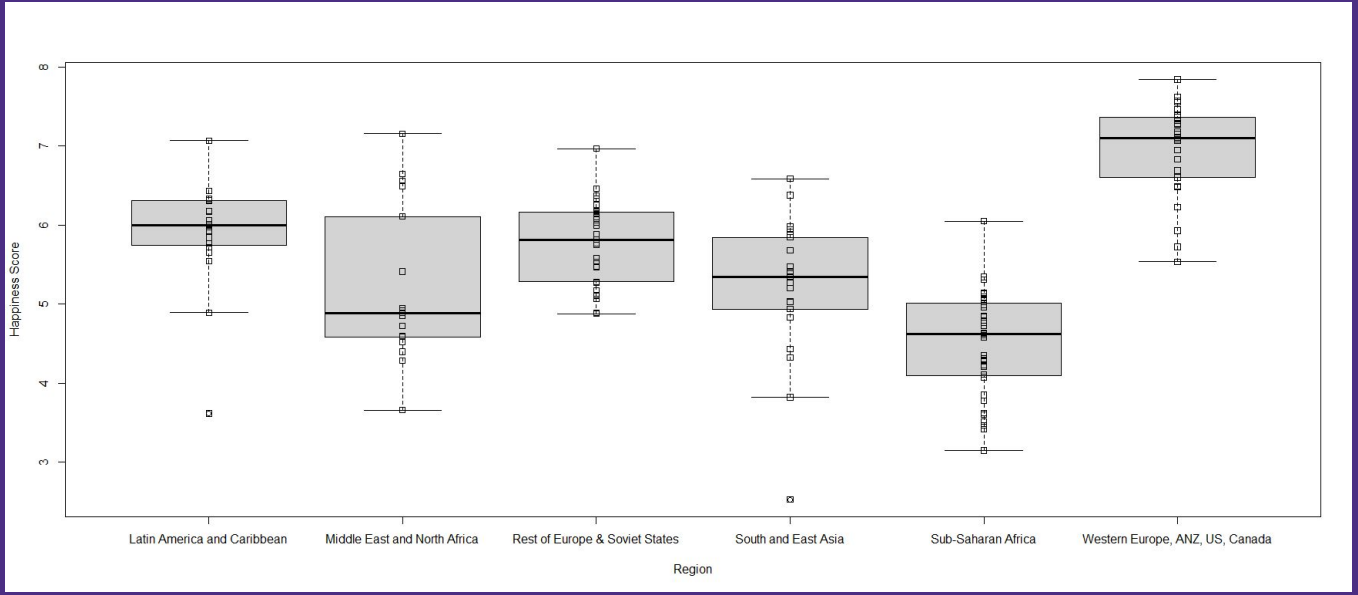
- Method Used: ANOVA test.

- Null Hypothesis: $H_0 : \mu_{R1} = \mu_{R2} = \mu_{R3} = \mu_{R4} = \mu_{R5} = \mu_{R6}$
  H_1: The means are not all equal

- Assumptions:
  - Independence
  - Equal variances
  - Large sample size or normal distribution

Are all regions of the world equally happy?

- Using ANOVA, we'd reject the null hypothesis with a p-value <2e-16
- The regions do not have constant variance, but a comparison of the equal variance t-test and Welch t-test showed little impact for violating that assumption.



Q&A and Discussion