

PSTAT 100 Final Project

By: Mason Delan, Drew Cockerham, Priyanka Pulikeshi, and Omar de Santiago

Data Description

Why was the data collected? What is the purpose of this dataset?

- The World Happiness Report is an annual publication that provides insights into the subjective well-being and happiness levels of countries around the world. The purpose of this dataset is to provide policymakers, politicians, and the general public with valuable information that contributes to a society's well-being. What are the observational units?

The observational units are a specific country during a particular year. What is the population of interest?

- The population of interest is the global population, and more specifically, the citizens of respective countries that are included in the data. This captures a wide range of perspectives and other nuances in order to represent happiness worldwide.

How was the sample obtained (e.g. random sampling, administrative data, convenience sampling, etc.)?

- The data regarding GDP per capita was obtained through the administrative data of particular countries. The rest of the sample was obtained through random sampling, specifically from data from the WHO and the Gallup World Poll. Every nation distributes surveys to its citizens in a way that guarantees a representative and varied sample. According to the Gallup World Poll website, the interviews cover at least 80% of a country's population and are representative of its non institutionalized adult population above the age of 15. This method guarantees that the results may be extrapolated to a wider population while also reducing bias.

Can inferences about the population be drawn from the sample?

- Inferences can be made about the sample population. The possibility that the sample is representative of the greater population rises when random sampling is used. But it's crucial to remember that poll results are arbitrary and impacted by

individual and cultural variances.

Dataset Description:

1. Country name: This column contains the names of different countries included in the report. Each row corresponds to a specific country.
2. Year: The year column represents the specific year for which the happiness data is recorded. It indicates the time period during which the survey or data collection took place.
3. Life Ladder: The Life Ladder column measures the overall subjective well-being or life satisfaction of individuals in a country. It is typically represented on a scale from 0 to 10, where higher values indicate greater life satisfaction.
4. Log GDP per capita: This column represents the logarithm of the Gross Domestic Product (GDP) per capita of a country, which is calculated by how much each country produces, divided by the number of people in the country. It serves as a measure of economic prosperity or standard of living, with higher values indicating higher GDP per person.
5. Social Support: The Social Support column assesses the availability and strength of social networks and support systems within a country. It measures the degree to which individuals have assistance from family, friends, and other social connections. The question associated with this column was "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?".
6. Health life expectancy: This column represents the average life expectancy or the number of years a person can expect to live in good health. It serves as an indicator of the physical health, mental health, and well-being of the population.
7. Freedom to make Life Choices: This column measures the extent to which individuals perceive having freedom and autonomy in making life choices, such as career, relationships, and personal decisions. Higher values indicate greater freedom. The question associated with this column is "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?".
8. Generosity: The Generosity column reflects the level of generosity or willingness to help others within a country's population. It measures the frequency of charitable donations and voluntary work. The question associated with this column is "Have you donated money to a charity in the past month?".

9. Perception of Corruption: This column captures the perceived level of corruption within a country. It assesses the extent to which corruption is perceived to be prevalent in public institutions and the overall trust in the government and public sector. The questions associated with this column are "Is corruption widespread throughout the government or not" and "Is corruption widespread within businesses or not?".
10. Positive affect: The Positive affect column measures the frequency and intensity of positive emotions experienced by individuals in a country. The result of this column is given by the average of yes or no answers based on emotions of laughter, enjoyment, and interest.
11. Negative affect: The Negative affect column assesses the frequency and intensity of negative emotions experienced by individuals in a country. The result of this column is given by the average of yes or no answers based on emotions of worry, sadness, and anger.

Data Analysis (Part 1)

```
In [2]: import numpy as np # linear algebra
import pandas as pd # data manipulation and processing
import altair as alt # data visualization
import statsmodels.formula.api as smf # regression model
import statsmodels.tools as sm # regression model
import warnings
warnings.filterwarnings('ignore')
```

```
In [3]: # import and clean up the World Happiness Report 2023 dataset
data = pd.read_csv('data/whr-2023.csv')
data.dropna()
data.head()

# compute pairwise correlations of variables
data_ = data.copy()
data_.drop(['Country name', 'year'], axis=1, inplace=True)
corr_data = data_.corr()
corr_data
```

```
Out[3]:
```

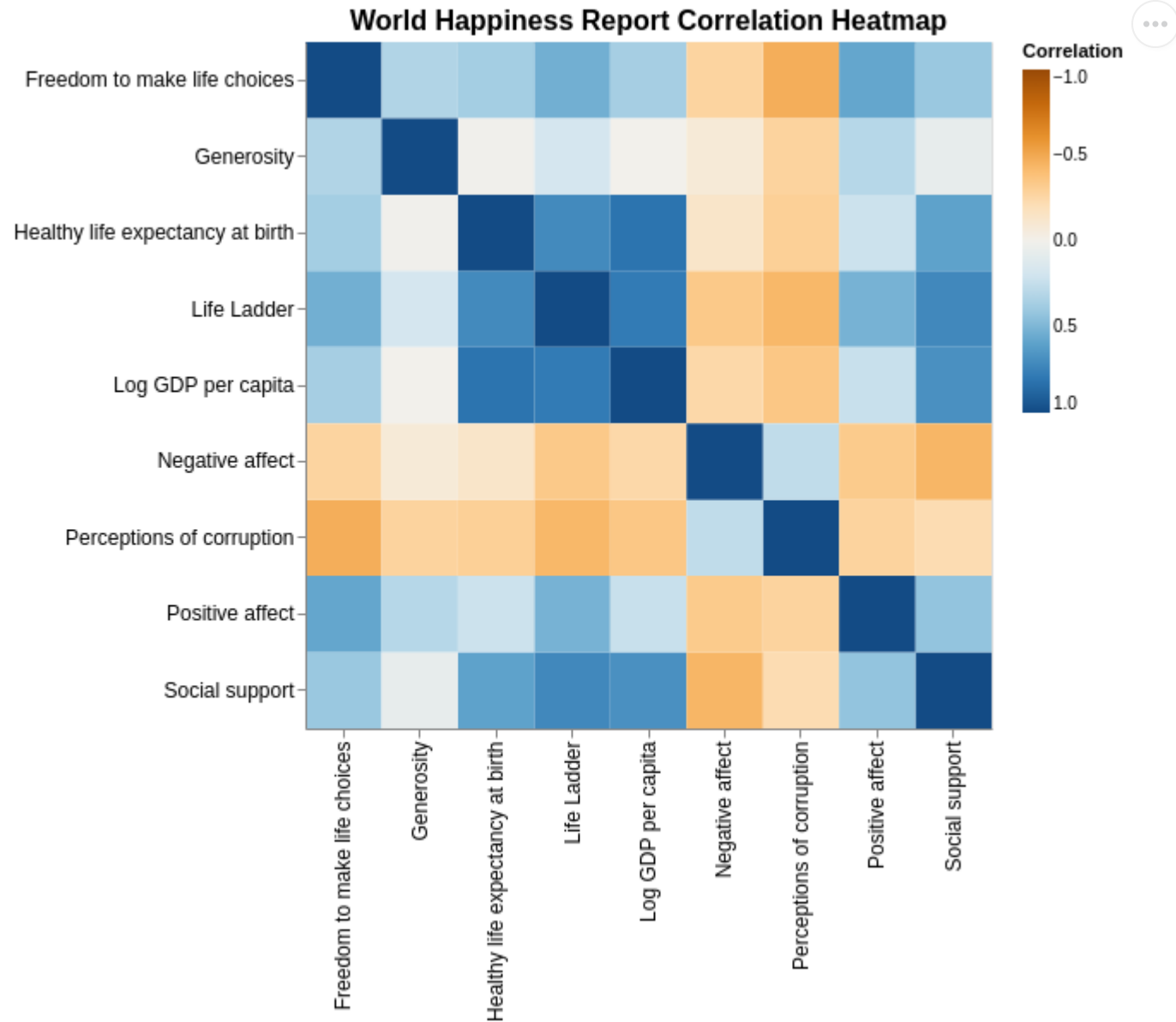
	Life Ladder	Log GDP per capita	Social support	Healthy life expectancy	Freedom to make life	Generosity	Perceptions of corruption	Positive affect	Negative affect
--	-------------	--------------------	----------------	-------------------------	----------------------	------------	---------------------------	-----------------	-----------------

	at birth				choices				
Life Ladder	1.000000	0.784868	0.721662	0.713499	0.534493	0.181630	-0.431500	0.518169	-0.339969
Log GDP per capita	0.784868	1.000000	0.683590	0.818126	0.367525	-0.000854	-0.352847	0.237933	-0.247541
Social support	0.721662	0.683590	1.000000	0.597659	0.409326	0.068572	-0.222551	0.431038	-0.441837
Healthy life expectancy at birth	0.713499	0.818126	0.597659	1.000000	0.373465	0.010775	-0.299016	0.223048	-0.140700
Freedom to make life choices	0.534493	0.367525	0.409326	0.373465	1.000000	0.325030	-0.476517	0.578680	-0.275438
Generosity	0.181630	-0.000854	0.068572	0.010775	0.325030	1.000000	-0.279435	0.307097	-0.080801
Perceptions of corruption	-0.431500	-0.352847	-0.222551	-0.299016	-0.476517	-0.279435	1.000000	-0.280606	0.266265
Positive affect	0.518169	0.237933	0.431038	0.223048	0.578680	0.307097	-0.280606	1.000000	-0.330301
Negative affect	-0.339969	-0.247541	-0.441837	-0.140700	-0.275438	-0.080801	0.266265	-0.330301	1.000000

```
In [4]: # create heatmap based on correlation matrix
corrlong1 = pd.melt(corr_data.reset_index(), id_vars='index', value_vars=corr_data.columns)
corrlong1.columns = ['row', 'col', 'Correlation']
fig1=alt.Chart(corrlong1).mark_rect().encode(
    x=alt.X('col:O', title=None),
    y=alt.Y('row:O', title=None),
    color=alt.Color('Correlation:Q', scale=alt.Scale(scheme='blueorange', domain=(1,-1))).properties(
        width=400,
        height=400,
        title='World Happiness Report Correlation Heatmap'
    ).configure_axis(
```

```
labelFontSize=12,  
)  
.configure_title(  
    fontSize=16  
)  
fig1
```

Out[4]:



We now have constructed a heatmap to determine possible variables of interest based on their respective correlations. In essence, we want to choose strong relationships to explore further. Some noticeably strong correlations include Log GDP per capita and Healthy life expectancy at birth. We want to further break down Log GDP per capita into two main groups, low and high, with the threshold being based on the median Log GDP of this dataset. This categorization will help us in further investigating the relationship.

```
In [5]: # create variable that indicates wealth (High or Low) based on Log GDP per capita
median_gdp = data['Log GDP per capita'].median()
data['Wealth'] = np.where(data['Log GDP per capita'] > median_gdp, 'High', 'Low')

# create variable that indicates time period (recession, covid, or other) based on Year
data['Indicator'] = data['year'].apply(lambda x: 'recession' if x in [2008, 2009, 2010]
                                     else 'covid' if x in [2020, 2021, 2022]
                                     else 'other'))

# drop Country name and year columns
data.drop(['Country name', 'year'], axis=1, inplace=True)
```

Based on our breakdown of the Log GDP per capita, we are interested in exploring the impacts of specific global events such as the Great Recession (2008-2009) & the COVID-19 Pandemic (2020-2022), and how these events may have impacted the happiness of populations in both high GDP nations and low GDP nations. By organizing the data further with these two categorical variables, we may see whether a change in these factors would have some significant impact on the relationships between various signifiers of happiness in nations across the world.

```
In [6]: # split data into 6 groups: high gdp during covid, high gdp during recession, high gdp other,
# low gdp during covid, low gdp during recession, and low gdp other
highgdp_rec = data[(data['Wealth'] == 'High') & (data['Indicator'] == 'recession')]
highgdp_rec.drop(['Wealth', 'Indicator'], axis=1, inplace=True)
lowgdp_rec = data[(data['Wealth'] == 'Low') & (data['Indicator'] == 'recession')]
lowgdp_rec.drop(['Wealth', 'Indicator'], axis=1, inplace=True)
highgdp_covid = data[(data['Wealth'] == 'High') & (data['Indicator'] == 'covid')]
highgdp_covid.drop(['Wealth', 'Indicator'], axis=1, inplace=True)
lowgdp_covid = data[(data['Wealth'] == 'Low') & (data['Indicator'] == 'covid')]
lowgdp_covid.drop(['Wealth', 'Indicator'], axis=1, inplace=True)
highgdp_other = data[(data['Wealth'] == 'High') & (data['Indicator'] == 'other')]
highgdp_other.drop(['Wealth', 'Indicator'], axis=1, inplace=True)
lowgdp_other = data[(data['Wealth'] == 'Low') & (data['Indicator'] == 'other')]
```

```

lowgdp_other.drop(['Wealth', 'Indicator'], axis=1, inplace=True)

# compute pairwise correlations of variables for each group
corr_mx1=highgdp_rec.corr()
corr_mx2=lowgdp_rec.corr()
corr_mx3=highgdp_covid.corr()
corr_mx4=lowgdp_covid.corr()
corr_mx5=highgdp_other.corr()
corr_mx6=lowgdp_other.corr()

# create heatmaps based on the correlation matrixes for each group
corr_mx_long1 = pd.melt(corr_mx1.reset_index(), id_vars='index', value_vars=corr_mx1.columns)
corr_mx_long1.columns = ['row', 'col', 'Correlation']

fig1=alt.Chart(corr_mx_long1).mark_rect().encode(
    x=alt.X('col:O', title=None),
    y=alt.Y('row:O', title=None),
    color=alt.Color('Correlation:Q', scale=alt.Scale(scheme='blueorange', domain=(1,-1)))
)
corr_mx_long2 = pd.melt(corr_mx2.reset_index(), id_vars='index', value_vars=corr_mx2.columns)
corr_mx_long2.columns = ['row', 'col', 'Correlation']

fig2=alt.Chart(corr_mx_long2).mark_rect().encode(
    x=alt.X('col:O', title=None),
    y=alt.Y('row:O', title=None),
    color=alt.Color('Correlation:Q', scale=alt.Scale(scheme='blueorange', domain=(1,-1)))
)
corr_mx_long3 = pd.melt(corr_mx3.reset_index(), id_vars='index', value_vars=corr_mx3.columns)
corr_mx_long3.columns = ['row', 'col', 'Correlation']

fig3=alt.Chart(corr_mx_long3).mark_rect().encode(
    x=alt.X('col:O', title=None),
    y=alt.Y('row:O', title=None),
    color=alt.Color('Correlation:Q', scale=alt.Scale(scheme='blueorange', domain=(1,-1)))
)
corr_mx_long4 = pd.melt(corr_mx4.reset_index(), id_vars='index', value_vars=corr_mx4.columns)
corr_mx_long4.columns = ['row', 'col', 'Correlation']

fig4=alt.Chart(corr_mx_long4).mark_rect().encode(
    x=alt.X('col:O', title=None),
    y=alt.Y('row:O', title=None),
    color=alt.Color('Correlation:Q', scale=alt.Scale(scheme='blueorange', domain=(1,-1)))
)

```

```

)
corr_mx_long5 = pd.melt(corr_mx5.reset_index(), id_vars='index', value_vars=corr_mx5.columns)
corr_mx_long5.columns = ['row', 'col', 'Correlation']

fig5=alt.Chart(corr_mx_long5).mark_rect().encode(
    x=alt.X('col:O', title=None),
    y=alt.Y('row:O', title=None),
    color=alt.Color('Correlation:Q', scale=alt.Scale(scheme='blueorange', domain=(1,-1)))
)

corr_mx_long6 = pd.melt(corr_mx6.reset_index(), id_vars='index', value_vars=corr_mx6.columns)
corr_mx_long6.columns = ['row', 'col', 'Correlation']

fig6=alt.Chart(corr_mx_long6).mark_rect().encode(
    x=alt.X('col:O', title=None),
    y=alt.Y('row:O', title=None),
    color=alt.Color('Correlation:Q', scale=alt.Scale(scheme='blueorange', domain=(1,-1)))
)

fig1 = fig1.properties(title='High GDP, recession',
    width=200,
    height=200,
)
fig2 = fig2.properties(title='Low GDP, recession',
    width=200,
    height=200,
)
fig3 = fig3.properties(title='High GDP, covid',
    width=200,
    height=200,
)
fig4 = fig4.properties(title='Low GDP, covid',
    width=200,
    height=200,
)
fig5 = fig5.properties(title='High GDP, other',
    width=200,
    height=200,
)
fig6 = fig6.properties(title='Low GDP, other',
    width=200,
    height=200,
)

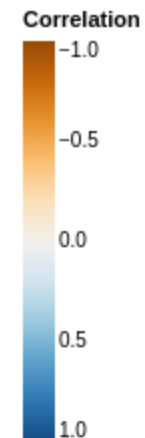
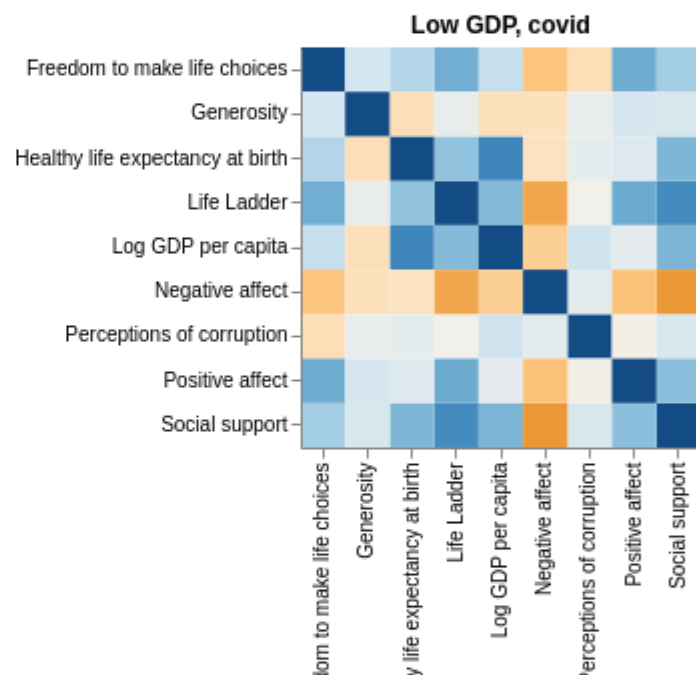
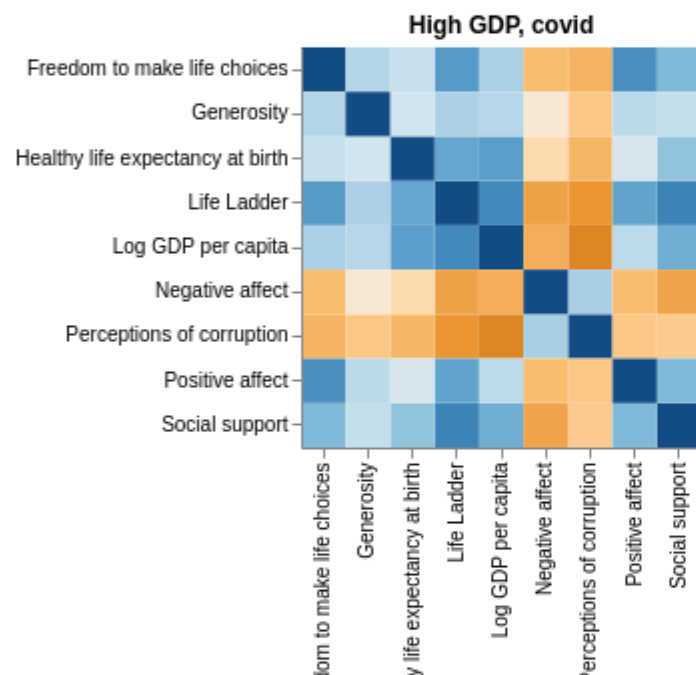
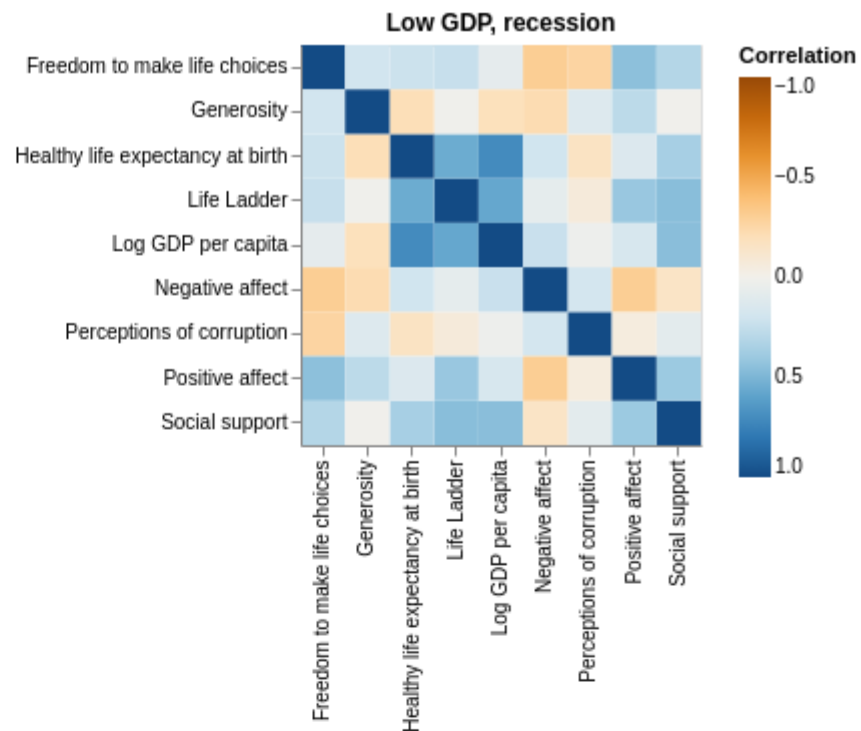
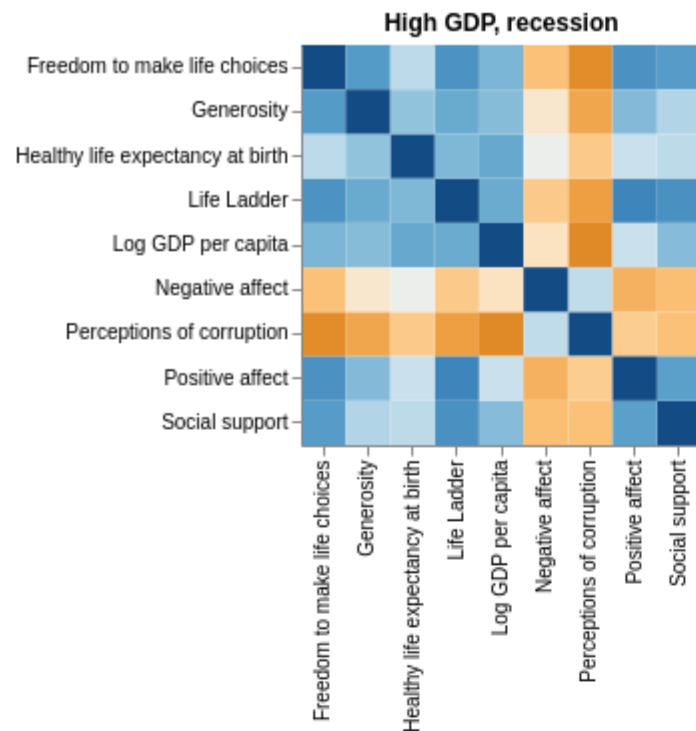
```

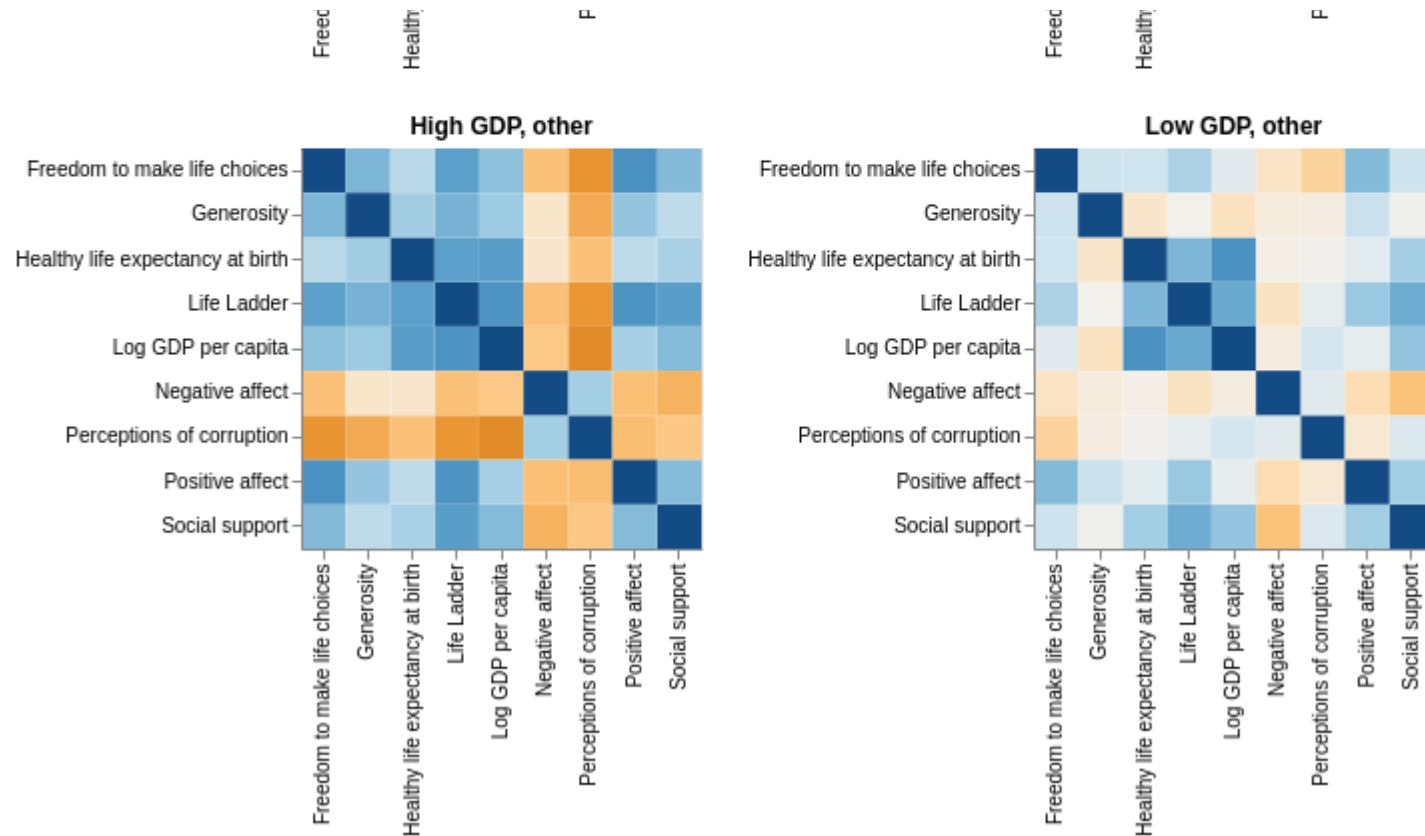


```
row1 = alt.hconcat(fig1, fig2)
row2 = alt.hconcat(fig3, fig4)
row3 = alt.hconcat(fig5, fig6)

final_layout = alt.vconcat(row1, row2, row3).configure_axis(
    labelFontSize=10,
).configure_title(
    fontSize=12
)

final_layout.display()
```





By plotting these heatmaps, we can visualize the variation in pairwise correlations between the following categories: High GDP & recession, Low GDP & recession, High GDP & covid, Low GDP & covid, High GDP & other, and Low GDP & other. The term 'other' will be our baseline for our correlations of interest, 'other' referring to time periods not included in the Great Recession or COVID-19 Pandemic. An interesting observation of the heatmaps is the discrepancy between High GDP & recession and Low GDP & recession with the relationship between Life Ladder and Social Support; this is something we will be exploring further

We were interested in answering the question of whether there was discrepancy in relationships between variables of interest when comparing data from high GDP nations to data from low GDP nations, as well as data from around the time of the 2008 financial crisis, COVID-19 pandemic, or neither. In the interest of exploratory analysis, we decided to create multiple linear regression models with Life Ladder, a variable that most comprehensively encapsulates the measure for happiness in this dataset, being the response variable. In terms of visualization, make a scatterplot of correlations of interest and then overlay the respective regression lines made with predicted values of the response for each of the 6 combinations of factors.

Question of Interest

How are the relationships between variables of interest, which are happiness signifiers, affected by variations in GDP among nations and the timing of data collection, specifically during the 2008 financial crisis or the COVID-19 pandemic?

Hypothesis: We have reason to believe that the differences in the characteristics of these two globally disruptive events would lead to discrepancies in the relationships between signifiers of happiness. For example, the unique way in which the COVID 19 pandemic affected people's social lives might result in a change in the relationship between social support and life ladder, as compared with during the recession or other years. Both events might see a reduction in mean life ladder scores, but the social support score might have a greater drop during COVID.

Data Analysis (Part 2)

Regression Model 1

$$\text{Life Ladder}_i = \beta_0 + \beta_1 \text{Social Support} + \underbrace{\beta_2 \text{High}_i}_{\text{GDP Level}} + \underbrace{\beta_3 \text{Recession}_i + \beta_4 \text{Covid}_i}_{\text{Time Period}} + \epsilon_i$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Other}) = \underbrace{\beta_0}_{\text{intercept}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Other}) = \underbrace{\beta_0 + \beta_2}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Recession}) = \underbrace{\beta_0 + \beta_3}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Recession}) = \underbrace{\beta_0 + \beta_2 + \beta_3}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Covid}) = \underbrace{\beta_0 + \beta_4}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder} \mid \text{High GDP, Covid}) = \underbrace{\beta_0 + \beta_2 + \beta_4}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

```
In [7]: data1 = data.copy()

# rename columns to prepare for applying regression
data1.rename(columns={"Life Ladder": "Life_Ladder"}, inplace = True)
data1.rename(columns={"Social support": "Social_Support"}, inplace = True)
data1.rename(columns={"Freedom to make life choices": "Freedom_to_make_life_choices"}, inplace = True)
data1.rename(columns={"Perceptions of corruption": "Perceptions_of_corruption"}, inplace = True)
data1.rename(columns={"Positive affect": "Positive_affect"}, inplace = True)
data1.rename(columns={"Negative affect": "Negative_affect"}, inplace = True)
data1.rename(columns={"Healthy life expectancy at birth": "Healthy_life_expectancy_at_birth"}, inplace = True)

# explanatory variable matrix with Social Support as quantitative variable and Wealth and Indicator as categorical
dataa1 = data1.copy()
density_encoded1 = pd.get_dummies(dataa1.Wealth, dtype=int)
density_encoded1_5 = pd.get_dummies(dataa1.Indicator, dtype=int)
x_vars = pd.concat([dataa1.Social_Support, density_encoded1, density_encoded1_5], axis = 1)
x = sm.tools.add_constant(x_vars)
x.head()
```

```
Out[7]:
```

	const	Social_Support	High	Low	covid	other	recession
0	1.0	0.451	0	1	0	0	1
1	1.0	0.552	0	1	0	0	1
2	1.0	0.539	0	1	0	0	1
3	1.0	0.521	0	1	0	1	0
4	1.0	0.521	0	1	0	1	0

```
In [8]: # apply regression using chosen model with Wealth = Low and Indicator = Other being the reference levels
fit1 = smf.ols(formula='Life_Ladder ~ Social_Support + C(Wealth, Treatment("Low")) + C(Indicator, Treatment("Other"))',
               data=data1).fit()
print(fit1.summary())
```

OLS Regression Results

```

=====
Dep. Variable:          Life_Ladder    R-squared:                0.604
Model:                  OLS           Adj. R-squared:           0.604
Method:                 Least Squares  F-statistic:              833.4
Date:                  Thu, 14 Dec 2023  Prob (F-statistic):       0.00
Time:                  00:00:50        Log-Likelihood:           -2345.9
No. Observations:      2186           AIC:                     4702.
Df Residuals:          2181           BIC:                     4730.
Df Model:              4
Covariance Type:       nonrobust
=====

```

```

=====
                                coef    std err          t      P>|t|      [0.025
-----
0.975]
-----
Intercept                    1.0950    0.116      9.423    0.000    0.867
1.323
C(Wealth, Treatment("Low"))[T.High]  0.7704    0.037    20.767    0.000    0.698
0.843
C(Indicator, Treatment("other"))[T.covid]  0.1839    0.042     4.373    0.000    0.101
0.266
C(Indicator, Treatment("other"))[T.recession]  0.0358    0.042     0.842    0.400   -0.047
0.119
Social_Support                4.8904    0.153    31.918    0.000    4.590
5.191
=====

```

```

=====
Omnibus:                  16.323    Durbin-Watson:           0.473
Prob(Omnibus):            0.000    Jarque-Bera (JB):        16.618
Skew:                    -0.210    Prob(JB):                0.000246
Kurtosis:                 2.922    Cond. No.                 18.0
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Here, we wondered if there was a way to improve the R^2 value in some way and decided to add an interaction term to the model that captured the effects of the Social Support, Wealth, and Indicator variables on eachother.

$$\begin{aligned}
\text{Life Ladder}_i = & \beta_0 + \beta_1 \text{Social Support} + \underbrace{\beta_2 \text{High}_i}_{\text{GDP Level}} + \underbrace{\beta_3 \text{Recession}_i + \beta_4 \text{Covid}_i}_{\text{Time Period}} \\
& + \underbrace{\beta_5 \text{Social Support}_i \text{Covid}_i + \beta_6 \text{Social Support}_i \text{Recession}_i}_{\text{Interactions}} \\
& + \underbrace{\beta_7 \text{Social Support}_i \text{High}_i \text{Covid}_i + \beta_8 \text{Social Support}_i \text{High}_i \text{Recession}_i + \beta_9 \text{Social Support}_i \text{High}_i \text{Other}_i}_{\text{Interactions}} + \epsilon_i
\end{aligned}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Other}) = \underbrace{\beta_0}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Other}) = \underbrace{\beta_0 + \beta_2 + \beta_9}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Recession}) = \underbrace{\beta_0 + \beta_3 + \beta_6}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Recession}) = \underbrace{\beta_0 + \beta_2 + \beta_3 + \beta_8}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Covid}) = \underbrace{\beta_0 + \beta_4 + \beta_5}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Covid}) = \underbrace{\beta_0 + \beta_2 + \beta_4 + \beta_7}_{\text{intercepts}} + \beta_1 \text{Social Support}$$

```

In [9]: # apply regression to above model including interaction term
fit2 = smf.ols(formula='''Life_Ladder ~ Social_Support + C(Wealth, Treatment("Low")) + C(Indicator, Treatment("Low")) + Social_Support:C(Wealth, Treatment("Low")):C(Indicator, Treatment("other"))''', data=data1).fit()
print(fit2.summary())

```

OLS Regression Results

```

=====
Dep. Variable:          Life_Ladder    R-squared:                0.638
Model:                  OLS           Adj. R-squared:          0.637
Method:                 Least Squares  F-statistic:             426.5
Date:                   Thu, 14 Dec 2023  Prob (F-statistic):      0.00

```

```

Time:                00:00:51   Log-Likelihood:        -2248.5
No. Observations:    2186      AIC:                4517.
Df Residuals:        2176      BIC:                4574.
Df Model:             9
Covariance Type:     nonrobust

```

					coef
std err	t	P> t	[0.025	0.975]	

Intercept					1.9707
0.152	12.953	0.000	1.672	2.269	
C(Wealth, Treatment("Low"))[T.High]					-3.1774
0.295	-10.754	0.000	-3.757	-2.598	
C(Indicator, Treatment("other"))[T.covid]					-0.7239
0.305	-2.374	0.018	-1.322	-0.126	
C(Indicator, Treatment("other"))[T.recession]					0.5519
0.311	1.777	0.076	-0.057	1.161	
Social_Support					3.7052
0.202	18.330	0.000	3.309	4.102	
Social_Support:C(Indicator, Treatment("other"))[T.covid]					1.3297
0.415	3.204	0.001	0.516	2.144	
Social_Support:C(Indicator, Treatment("other"))[T.recession]					-0.7160
0.409	-1.749	0.080	-1.519	0.087	
Social_Support:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[covid]					4.2995
0.361	11.903	0.000	3.591	5.008	
Social_Support:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[other]					4.6819
0.347	13.474	0.000	4.000	5.363	
Social_Support:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[recession]					4.8485
0.355	13.646	0.000	4.152	5.545	
=====					
Omnibus:		21.066	Durbin-Watson:		0.521
Prob(Omnibus):		0.000	Jarque-Bera (JB):		21.564
Skew:		-0.243	Prob(JB):		2.08e-05
Kurtosis:		2.964	Cond. No.		68.9
=====					

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

After adding the interaction term, the R^2 of our model improved slightly from 0.604 to 0.638. This means that around 63.8% of the variance in the model can be explained by the independent variables.

Fitted Model:

$$\begin{aligned}\widehat{\text{Life Ladder}} = & 1.9707 + 3.7052(\text{Social Support}) - 3.1774(\text{High}) + 0.5519(\text{Recession}) \\ & - 0.7239(\text{Covid}) + 1.3297(\text{Social Support} * \text{Covid}) - 0.7160(\text{Social Support} * \text{Recession}) \\ & + 4.6819(\text{Social Support} * \text{High} * \text{Other}) + 4.2995(\text{Social Support} * \text{High} * \text{Covid}) \\ & + 4.8485(\text{Social Support} * \text{High} * \text{Recession})\end{aligned}$$

Low GDP and Other:

$$\widehat{\text{Life Ladder}} = 1.9707 + 3.7052(\text{Social Support})$$

The expected Life Ladder is the sum of the intercept β_0 and the impact of a one-unit increase in Social Support.

High GDP and Other:

$$\widehat{\text{Life Ladder}} = 3.4752 + 3.7052(\text{Social Support})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Social Support, the impact of being in the High GDP category, and the interaction between the conditions Social Support, High GDP, and other.

Low GDP and Covid:

$$\widehat{\text{Life Ladder}} = 2.5765 + 3.7052(\text{Social Support})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Social Support, the impact of Covid, and the interaction between the conditions Social Support and Covid.

High GDP and Covid:

$$\widehat{\text{Life Ladder}} = 2.3689 + 3.7052(\text{Social Support})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Social Support, the impact of being in the High GDP category and of Covid, and the interaction between the conditions Social Support, High GDP, and Covid.

Low GDP and Recession:

$$\text{Life Ladder} = 1.8066 + 3.7052(\text{Social Support})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Social Support, and the impact of the Recession, and the interaction between the conditions Social Support, High GDP, Recession.

High GDP and Recession:

$$\text{Life Ladder} = 4.1937 + 3.7052(\text{Social Support})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Social Support, the impact of being in the High GDP category and of the Recession, and the interaction between the conditions Social Support, High GDP, and Recession.

Interpretation:

The intercepts highlight the impact of the two categorical variables, Wealth and Indicator, and the interaction variables on the Life Ladder of countries having the same Social Support. Based on these equations, the highest intercept occurs in the "High GDP and Recession" condition at 4.1937 and the lowest intercept occurs in the "Low GDP and Recession" condition at 1.8066. For each time period except the Covid, the conditions with High GDP have a higher intercept than their Low GDP counterparts. In order to better visualize these discrepancies, we decided to plot the Social Support and Life Ladder as a base scatterplot and layered each of the six regression lines on top. We used the observed Social Support values from the dataset to calculate predicted response values for the regression lines.

```
In [10]: reg_data1 = data1.copy()

# make a list of sample coefficients from model fit
coefficients = {
    'Intercept': 1.9707,
    'C(Wealth, Treatment("Low"))[T.High]': -3.1774,
    'C(Indicator, Treatment("other"))[T.covid]': -0.7239,
    'C(Indicator, Treatment("other"))[T.recession]': 0.5519,
    'Social_Support': 3.7052,
    'Social_Support:C(Indicator, Treatment("other"))[T.covid]': 1.3297,
    'Social_Support:C(Indicator, Treatment("other"))[T.recession]': -0.5519,
    'Social_Support:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[covid]': -0.7239,
    'Social_Support:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[other]': 0.5519
```

```

    'Social_Support:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[recession]': 4
}

# calculate predicted values of Life Ladder based on the estimated coefficients plus specific interactions
reg_data1['predicted_ll_lo'] = (coefficients['Intercept'] + (coefficients['Social_Support'] * reg_data1['Sc
reg_data1['predicted_ll_lc_i'] = (coefficients['Intercept'] + coefficients['C(Indicator, Treatment("other")
(coefficients['Social_Support'] * reg_data1['Social_Support']) + coefficients['Social_Support:C(Indicator,
reg_data1['predicted_ll_lr_i'] = (coefficients['Intercept'] + coefficients['C(Indicator, Treatment("other")
(coefficients['Social_Support'] * reg_data1['Social_Support']) + coefficients['Social_Support:C(Indicator,
reg_data1['predicted_ll_ho_i'] = (coefficients['Intercept'] + coefficients['C(Wealth, Treatment("Low"))[T.H
(coefficients['Social_Support'] * reg_data1['Social_Support']) + coefficients['Social_Support:C(Wealth, Tre
reg_data1['predicted_ll_hc_i'] = (coefficients['Intercept'] + coefficients['C(Wealth, Treatment("Low"))[T.H
(coefficients['Social_Support'] * reg_data1['Social_Support']) + coefficients['Social_Support:C(Wealth, Tre
reg_data1['predicted_ll_hr_i'] = (coefficients['Intercept'] + coefficients['C(Wealth, Treatment("Low"))[T.H
(coefficients['Social_Support'] * reg_data1['Social_Support']) + coefficients['Social_Support:C(Wealth, Tre

# create scatter plot of Social Support against Life Ladder
chart1 = alt.Chart(reg_data1).mark_circle(color='pink').encode(
    x=alt.X('Social_Support:Q', title='Social Support'),
    y='Life_Ladder:Q',
)

# create regression lines based on predicted values from each group
line_ll_lo = alt.Chart(reg_data1).mark_line(color='coral').encode(
    x='Social_Support:Q',
    y='predicted_ll_lo:Q',
)
line_ll_lc_i = alt.Chart(reg_data1).mark_line(color='green').encode(
    x='Social_Support:Q',
    y='predicted_ll_lc_i:Q'
)
line_ll_lr_i = alt.Chart(reg_data1).mark_line(color='red').encode(
    x='Social_Support:Q',
    y='predicted_ll_lr_i:Q'
)
line_ll_ho_i = alt.Chart(reg_data1).mark_line(color='purple').encode(
    x='Social_Support:Q',
    y='predicted_ll_ho_i:Q'
)
line_ll_hc_i = alt.Chart(reg_data1).mark_line(color='black').encode(
    x='Social_Support:Q',

```

```

        y='predicted_ll_hc_i:Q'
    )
    line_ll_hr_i = alt.Chart(reg_data1).mark_line(color='blue').encode(
        x='Social_Support:Q',
        y='predicted_ll_hr_i:Q'
    )
    title_ll_lo = alt.Chart({'values': [{'x': 0.015, 'y': 2.7, 'title': 'Low, Other'}]}).mark_text(
        align='left',
        baseline='middle',
        dx=5,
        color='coral',
        fontSize=11
    ).encode(
        x='x:Q',
        y='y:Q',
        text='title:N'
    )
    title_ll_lc_i = alt.Chart({'values': [{'x': 0.015, 'y': 3.45, 'title': 'Low, Covid, Interaction'}]}).mark_t
        align='left',
        baseline='middle',
        dx=5,
        color='green',
        fontSize=11
    ).encode(
        x='x:Q',
        y='y:Q',
        text='title:N'
    )
    title_ll_lr_i = alt.Chart({'values': [{'x': 0.015, 'y': 2.45, 'title': 'Low, Recession, Interaction'}]}).m
        align='left',
        baseline='middle',
        dx=5,
        color='red',
        fontSize=11
    ).encode(
        x='x:Q',
        y='y:Q',
        text='title:N'
    )
    title_ll_ho_i = alt.Chart({'values': [{'x': 0.015, 'y': 4.25, 'title': 'High, Other, Interaction'}]}).mark
        align='left',
        baseline='middle',

```

```

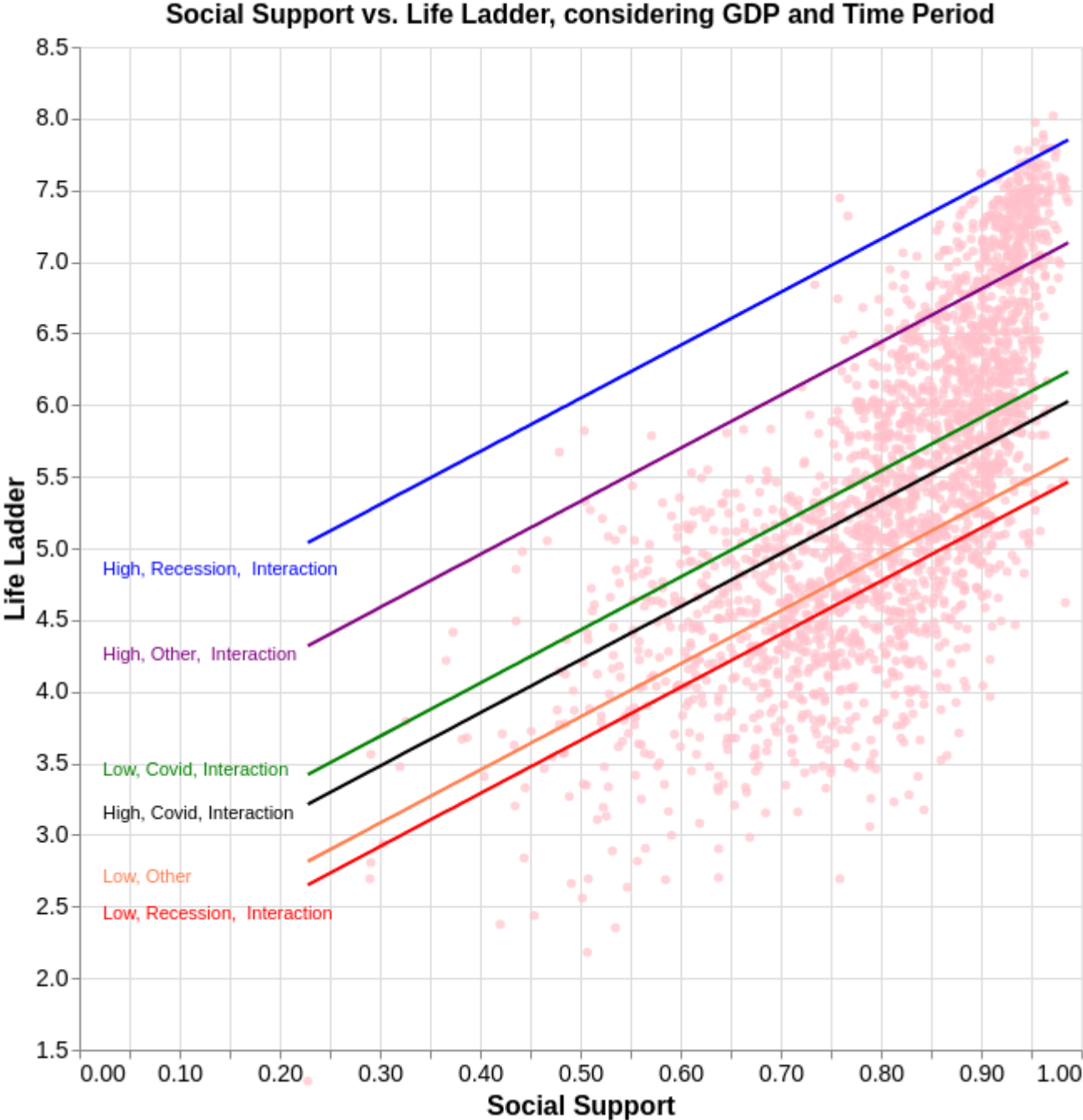
        dx=5,
        color='purple',
        fontSize=11
    ).encode(
        x='x:Q',
        y='y:Q',
        text='title:N'
    )
title_ll_hc_i = alt.Chart({'values': [{'x': 0.015, 'y': 3.15, 'title': 'High, Covid, Interaction'}]}).mark_
    align='left',
    baseline='middle',
    dx=5,
    color='black',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)
title_ll_hr_i = alt.Chart({'values': [{'x': 0.015, 'y': 4.85, 'title': 'High, Recession, Interaction'}]}).
    align='left',
    baseline='middle',
    dx=5,
    color='blue',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)

# combine scatterplot and regression lines
final_chart1 = (chart1 + line_ll_lo + line_ll_lc_i + line_ll_lr_i + line_ll_ho_i + line_ll_hc_i + line_ll_hr_i
    + title_ll_ho_i + title_ll_hc_i + title_ll_hr_i).properties(
    width=600,
    height=600,
    title='Social Support vs. Life Ladder, considering GDP and Time Period'
).configure_axis(
    labelFontSize=14,
    titleFontSize=16,
).configure_title(
    fontSize=16

```

```
).encode(  
    y=alt.Y(title='Life Ladder', scale=alt.Scale(domain=[1.5, 8.5]))  
)  
final_chart1
```

Out[10]:



By the graph, we can see that all of the lines representing each variation of the model have the same slope, but different intercepts. According to this regression model, countries having the same average social support have a higher average life ladder score if they are a country with high GDP compared to if they are a country with low GDP, except in the case of the Covid time period. Low GDP countries having the same average social support have a higher average Life Ladder score during Covid as compared to during the Recession or other years. There is a significant absolute difference between the High GDP/Recession & Low GDP/Recession intercepts of about 2.3871, presenting us with another very interesting point of contention, why is this discrepancy far greater than our baseline 'other'? Along the same lines, why is the discrepancy between High GDP/Covid & Low GDP/Covid intercepts considerably smaller than than our baseline 'other'?

Regression Model 2

We decided to further investigate the impact of the discrepancies in the High and Low GDP countries in the Recession and Covid time periods by creating a new model with Social Support replaced by Freedom to Make Life Choices. In this way, we would be measuring the same response, Life Ladder, and seeing if the impact of the categorical variables and their interaction with a separate quantitative variable had a similar affect as it did with our first model.

$$\text{Life Ladder}_i = \beta_0 + \beta_1 \text{Freedom to Make Life Choices} + \underbrace{\beta_2 \text{High}_i}_{\text{GDP Level}} + \underbrace{\beta_3 \text{Recession}_i + \beta_4 \text{Covid}_i}_{\text{Time Period}} + \epsilon_i$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Other}) = \underbrace{\beta_0}_{\text{intercept}} + \beta_1 \text{Freedom to Make Life Choices}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Other}) = \underbrace{\beta_0 + \beta_2}_{\text{intercepts}} + \beta_1 \text{Freedom to Make Life Choices}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Recession}) = \underbrace{\beta_0 + \beta_3}_{\text{intercepts}} + \beta_1 \text{Freedom to Make Life Choices}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Recession}) = \underbrace{\beta_0 + \beta_2 + \beta_3}_{\text{intercepts}} + \beta_1 \text{Freedom to Make Life Choices}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{Low GDP, Covid}) = \underbrace{\beta_0 + \beta_4}_{\text{intercepts}} + \beta_1 \text{Freedom to Make Life Choices}$$

$$\mathbb{E}(\text{Life Ladder}) \mid \text{High GDP, Covid}) = \underbrace{\beta_0 + \beta_2 + \beta_4}_{\text{intercepts}} + \beta_1 \text{Freedom to Make Life Choices}$$

```
In [11]: data2 = data1.copy()

# explanatory variable matrix with Freedom to make Life Choices as quantitative variable and Wealth and Indicator as categorical variables
dataa2 = data2.copy()
density_encoded2 = pd.get_dummies(dataa1.Wealth, dtype=int)
density_encoded2_5 = pd.get_dummies(dataa1.Indicator, dtype=int)
x_vars2 = pd.concat([dataa2.Freedom_to_make_life_choices, density_encoded2, density_encoded2_5], axis = 1)
x2 = sm.tools.add_constant(x_vars2)
x2.head()
```

```
Out[11]:
```

	const	Freedom_to_make_life_choices	High	Low	covid	other	recession
0	1.0	0.718	0	1	0	0	1
1	1.0	0.679	0	1	0	0	1
2	1.0	0.600	0	1	0	0	1
3	1.0	0.496	0	1	0	1	0
4	1.0	0.531	0	1	0	1	0

```
In [12]: # apply regression to above model including interaction variable
fit2 = smf.ols(formula='''Life_Ladder ~ Freedom_to_make_life_choices + C(Wealth, Treatment("Low")) + C(Indicator, Treatment("other"))''', data=data2).fit()
print(fit2.summary())
```

```

                        OLS Regression Results
=====
Dep. Variable:          Life_Ladder      R-squared:                0.582
Model:                  OLS              Adj. R-squared:          0.580
Method:                 Least Squares     F-statistic:            333.3
Date:                  Thu, 14 Dec 2023   Prob (F-statistic):      0.00
Time:                  00:00:51           Log-Likelihood:         -2393.2
No. Observations:      2166              AIC:                   4806.
Df Residuals:          2156              BIC:                   4863.

```

Df Model: 9
Covariance Type: nonrobust

=====					
coef	std err	t	P> t	[0.025	0.975]

Intercept					
3.3119	0.130	25.524	0.000	3.057	3.566
C(Wealth, Treatment("Low"))[T.High]					
-0.5854	0.183	-3.196	0.001	-0.945	-0.226
C(Indicator, Treatment("other"))[T.covid]					
-1.3826	0.311	-4.452	0.000	-1.992	-0.774
C(Indicator, Treatment("other"))[T.recession]					
0.5247	0.220	2.382	0.017	0.093	0.957
Freedom_to_make_life_choices					
1.9841	0.180	11.013	0.000	1.631	2.337
Freedom_to_make_life_choices:C(Indicator, Treatment("other"))[T.covid]					
1.8772	0.409	4.591	0.000	1.075	2.679
Freedom_to_make_life_choices:C(Indicator, Treatment("other"))[T.recession]					
-0.5785	0.328	-1.764	0.078	-1.222	0.065
Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[covid]					
2.1542	0.247	8.712	0.000	1.669	2.639
Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[other]					
2.3936	0.241	9.939	0.000	1.921	2.866
Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[recession]					
n]	2.6484	0.275	9.644	0.000	2.110 3.187
=====					
Omnibus:		55.870	Durbin-Watson:		0.556
Prob(Omnibus):		0.000	Jarque-Bera (JB):		60.210
Skew:		-0.383	Prob(JB):		8.43e-14
Kurtosis:		3.285	Cond. No.		48.2
=====					

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fitted Model:

$$\text{Life Ladder} = 3.3119 + 1.9841(\text{Freedom to Make Life Choices}) - 0.5854(\text{High}) + 0.5247(\text{Recession}) - 1.3826(\text{Covid})$$

$$\begin{aligned}
&+1.8772(\text{Freedom to Make Life Choices}*\text{Covid}) - 0.5785(\text{Freedom to Make Life Choices}*\text{Recession}) \\
&+2.3936(\text{Freedom to Make Life Choices}*\text{High}*\text{Other}) + 2.1542(\text{Freedom to Make Life Choices}*\text{High}*\text{Covid}) \\
&+2.6484(\text{Freedom to Make Life Choices}*\text{High}*\text{Recession})
\end{aligned}$$

Low GDP and Other:

$$\text{Life Ladder} = 3.3119 + 1.9841(\text{Freedom to Make Life Choices})$$

The expected Life Ladder is the sum of the intercept β_0 and the impact of a one-unit increase in Freedom to Make Life Choices.

High GDP and Other:

$$\text{Life Ladder} = 5.1201 + 1.9841(\text{Freedom to Make Life Choices})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Freedom to Make Life Choices, the impact of being in the High GDP category, and the interaction between the conditions Freedom to Make Life Choices, High GDP, and other.

Low GDP and Covid:

$$\text{Life Ladder} = 3.8065 + 1.9841(\text{Freedom to Make Life Choices})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Freedom to Make Life Choices, the impact of Covid, and the interaction between the conditions Freedom to Make Life Choices and Covid.

High GDP and Covid:

$$\text{Life Ladder} = 3.4981 + 1.9841(\text{Freedom to Make Life Choices})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Freedom to Make Life Choices, the impact of being in the High GDP category and of Covid, and the interaction between the conditions Freedom to Make Life Choices, High GDP, and Covid.

Low GDP and Recession:

$$\text{Life Ladder} = 3.2581 + 1.9841(\text{Freedom to Make Life Choices})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Freedom to Make Life Choices, and the impact of the Recession, and the interaction between the conditions Freedom to Make Life Choices, High GDP, and Recession.

High GDP and Recession:

$$\text{Life Ladder} = 5.8996 + 1.9841(\text{Freedom to Make Life Choices})$$

The expected Life Ladder is the sum of the intercept β_0 , the impact of a one-unit increase in Freedom to Make Life Choices, the impact of being in the High GDP category and of the Recession, and the interaction between the conditions Freedom to Make Life Choices, High GDP, and Recession.

Interpretation:

The intercepts highlight the impact of the two categorical variables, Wealth and Indicator, and the interaction variables on the Life Ladder of countries having the same Freedom to Make Life Choices. Based on these equations, the highest intercept occurs in the "High GDP and Recession" condition at 5.8996 and the lowest intercept occurs in the "Low GDP and Recession" condition at 3.2581. Notably, the "High GDP and Recession" and "Low GDP and Recession" had the highest and lowest intercepts, respectively, in our first model involving Social Support as well. Again, as was the case with the first model, for each time period except the Covid, the conditions with High GDP have a higher intercept than their Low GDP counterparts. In order to better visualize these discrepancies, we decided to plot the Freedom to Make Life Choices and Life Ladder as a base scatterplot and layered each of the six regression lines on top. We used the observed Freedom to Make Life Choices values from the dataset to calculate predicted response values for the regression lines.

```
In [13]: reg_data2 = data2.copy()

# make a list of sample coefficients from model fit
coefficients2 = {
    'Intercept': 3.3119,
    'C(Wealth, Treatment("Low"))[T.High]': -0.5854,
    'C(Indicator, Treatment("other"))[T.covid]': -1.3826,
    'C(Indicator, Treatment("other"))[T.recession]': 0.5247,
    'Freedom_to_make_life_choices': 1.9841,
    'Freedom_to_make_life_choices:C(Indicator, Treatment("other"))[T.covid]': 1.8772,
    'Freedom_to_make_life_choices:C(Indicator, Treatment("other"))[T.recession]': -0.5785,
    'Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[covi
    'Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[othe
```

```

    'Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("other"))[rece
}

# calculate predicted values of Life Ladder based on the estimated coefficients plus specific interactions
reg_data2['predicted_ll_lo2'] = (coefficients2['Intercept'] + (coefficients2['Freedom_to_make_life_choices'
reg_data2['predicted_ll_lc_i2'] = (coefficients2['Intercept'] + coefficients2['C(Indicator, Treatment("othe
(coefficients2['Freedom_to_make_life_choices'] * reg_data2['Freedom_to_make_life_choices']) + coefficients2
reg_data2['predicted_ll_lr_i2'] = (coefficients2['Intercept'] + coefficients2['C(Indicator, Treatment("othe
(coefficients2['Freedom_to_make_life_choices'] * reg_data2['Freedom_to_make_life_choices']) + coefficients2
reg_data2['predicted_ll_ho_i2'] = (coefficients2['Intercept'] + coefficients2['C(Wealth, Treatment("Low"))[
(coefficients2['Freedom_to_make_life_choices'] * reg_data2['Freedom_to_make_life_choices']) +
coefficients2['Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("oth
reg_data2['predicted_ll_hc_i2'] = (coefficients2['Intercept'] + coefficients2['C(Wealth, Treatment("Low"))[
(coefficients2['Freedom_to_make_life_choices'] * reg_data2['Freedom_to_make_life_choices']) +
coefficients2['Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("oth
reg_data2['predicted_ll_hr_i2'] = (coefficients2['Intercept'] + coefficients2['C(Wealth, Treatment("Low"))[
(coefficients2['Freedom_to_make_life_choices'] * reg_data2['Freedom_to_make_life_choices']) +
coefficients2['Freedom_to_make_life_choices:C(Wealth, Treatment("Low"))[T.High]:C(Indicator, Treatment("oth

# create scatter plot of Social Support against Life Ladder
chart2 = alt.Chart(reg_data2).mark_circle(color='pink').encode(
    x=alt.X('Freedom_to_make_life_choices:Q', title='Freedom to Make Life Choices'),
    y='Life_Ladder:Q',
)

# create regression lines based on predicted values from each group
line_ll_lo2 = alt.Chart(reg_data2).mark_line(color='coral').encode(
    x='Freedom_to_make_life_choices:Q',
    y='predicted_ll_lo2:Q',
)
line_ll_lc_i2 = alt.Chart(reg_data2).mark_line(color='green').encode(
    x='Freedom_to_make_life_choices:Q',
    y='predicted_ll_lc_i2:Q'
)
line_ll_lr_i2 = alt.Chart(reg_data2).mark_line(color='red').encode(
    x='Freedom_to_make_life_choices:Q',
    y='predicted_ll_lr_i2:Q'
)
line_ll_ho_i2 = alt.Chart(reg_data2).mark_line(color='purple').encode(
    x='Freedom_to_make_life_choices:Q',
    y='predicted_ll_ho_i2:Q'

```

```

)
line_ll_hc_i2 = alt.Chart(reg_data2).mark_line(color='black').encode(
    x='Freedom_to_make_life_choices:Q',
    y='predicted_ll_hc_i2:Q'
)
line_ll_hr_i2 = alt.Chart(reg_data2).mark_line(color='blue').encode(
    x='Freedom_to_make_life_choices:Q',
    y='predicted_ll_hr_i2:Q'
)
title_ll_lo2 = alt.Chart({'values': [{'x': 0.015, 'y': 3.75, 'title': 'Low, Other'}]}).mark_text(
    align='left',
    baseline='middle',
    dx=5,
    color='coral',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)
title_ll_lc_i2 = alt.Chart({'values': [{'x': 0.015, 'y': 4.35, 'title': 'Low, Covid, Interaction'}]}).mark_
    align='left',
    baseline='middle',
    dx=5,
    color='green',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)
title_ll_lr_i2 = alt.Chart({'values': [{'x': 0.015, 'y': 3.5, 'title': 'Low, Recession, Interaction'}]}).m
    align='left',
    baseline='middle',
    dx=5,
    color='red',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)

```

```

title_ll_ho_i2 = alt.Chart({'values': [{'x': 0.015, 'y': 5.45, 'title': 'High, Other, Interaction'}]}).mar
    align='left',
    baseline='middle',
    dx=5,
    color='purple',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)
title_ll_hc_i2 = alt.Chart({'values': [{'x': 0.015, 'y': 4, 'title': 'High, Covid, Interaction'}]}).mark_te
    align='left',
    baseline='middle',
    dx=5,
    color='black',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)
title_ll_hr_i2 = alt.Chart({'values': [{'x': 0.015, 'y': 6.3, 'title': 'High, Recession, Interaction'}]}).
    align='left',
    baseline='middle',
    dx=5,
    color='blue',
    fontSize=11
).encode(
    x='x:Q',
    y='y:Q',
    text='title:N'
)

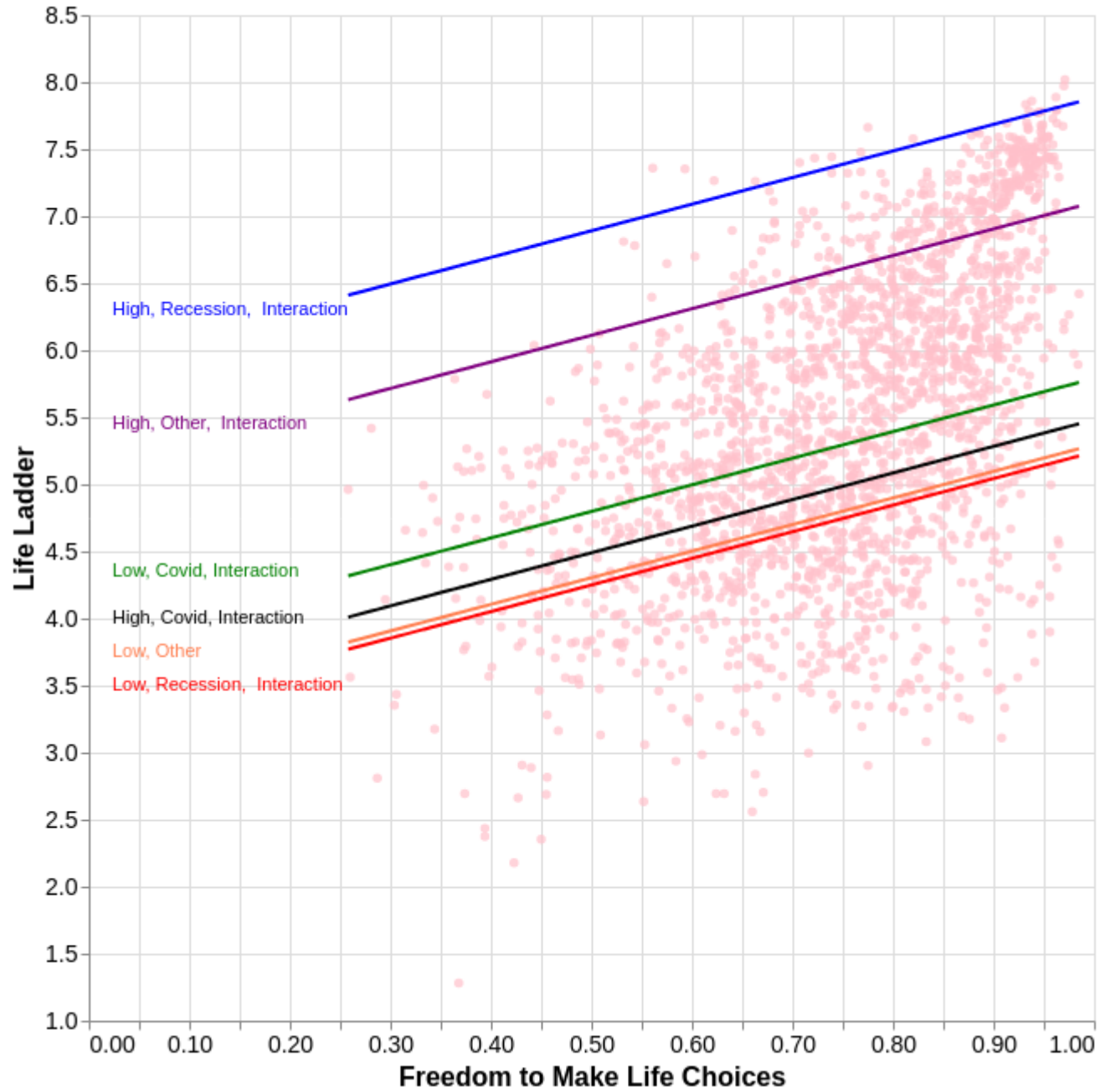
# combine scatterplot and regression lines
final_chart2 = (chart2 + line_ll_lo2 + line_ll_lc_i2 + line_ll_lr_i2 + line_ll_ho_i2 + line_ll_hc_i2 + line
    title_ll_lc_i2 + title_ll_lr_i2 + title_ll_ho_i2 + title_ll_hc_i2 + title_ll_hr_i2).properties(
    width=600,
    height=600,
    title='Freedom to Make Life Choices vs. Life Ladder, considering GDP and Time Period'
).configure_axis(
    labelFontSize=14,

```

```
    titleFontSize=16,  
) .configure_title(  
    fontSize=16  
) .encode(  
    y=alt.Y(title='Life Ladder', scale=alt.Scale(domain=[1.0, 8.5]))  
)  
final_chart2
```


Out[13]:

Freedom to Make Life Choices vs. Life Ladder, considering GDP and Time Period



According to this regression model, countries having the same average Freedom to Make Life Choices have a higher average life ladder score if they are a country with high GDP compared to if they are a country with low GDP, except in the case of the Covid time period. Low GDP countries having the same average Freedom to Make Life Choices have a higher average Life Ladder score during Covid as compared to during the Recession or other years. The largest intercept discrepancy based on GDP occurs during the Recession, with the absolute difference between the High GDP/Recession & Low GDP/Recession intercepts being 2.6415. The smallest intercept discrepancy based on GDP occurs during Covid, with the absolute difference between the High GDP/Covid & Low GDP/Covid intercepts being a mere 0.3084.

Interestingly, the patterns observed above were reflected, with different specific values of course, in the first model that used Social Support as a quantitative variable instead of Freedom to Make Life Choices. This indicates that there is some notable consistency as to the impact of the various combinations of the Wealth and Indicator variables on the relationship between various variables and Life Ladder.

Summary

- Based on the heatmaps of the 6 conditions, Covid had a statistically significant influence on 'Healthy Life Expectancy at Birth' between High and Low GDP nations.
- From both the first and second regression models, the individual impact of each condition remains relatively consistent in terms of intercepts. Intercepts of various condition combinations from highest to lowest in both models:
 1. High GDP and Recession
 2. High GDP and Other
 3. Low GDP and Covid
 4. High GDP and Covid
 5. Low GDP and Other
 6. Low GDP and Recession
- The difference between Low and High GDP intercepts is minimized during COVID-19 while the Recession saw a higher intercept difference.
- There is a clear positive correlation between Social Support and Life Ladder, while there is a weaker positive correlation between Social Support and Freedom to make life choices.
 - Lowest intercept from condition regression lines is higher in the second model than the first model.

- Countries that have the same average Social Support have a higher average Life Ladder score if they are a country with high GDP compared to if they are a country with low GDP, except in the case of the Covid time period.
- In the Freedom to make life choices vs Life Ladder scatterplot, the regression lines for low/covid, high/covid, low/other, low/recession all are very similar to one another. The regression lines representing High/other is relatively high and high/recession is even higher. Interpretation: Given the same 'Freedom to make life choices' score, high GDP nations on average would have lower life ladder scores during COVID than during other years.
- The relationship between 'Freedom to make life choices' and 'Life Ladder' is the same for the average low GDP nation whether in the 'recession' or 'other' bin, however it is slightly changed during COVID. Given the same 'Freedom to make life choices' score a low GDP nation would have a slightly elevated 'life ladder' score.
- In both models, the largest intercept discrepancy based on GDP occurs during the Recession, with the absolute difference between the High GDP/Recession & Low GDP/Recession intercepts being 2.3871 and 2.6415 in Model 1 and 2, respectively.
- For both models, interactions between Low GDP countries during the Recession and Social Support/Freedom to make life Choices had similarly high p-values. This may imply that low GDP countries in general were less affected by the Recession than high GDP countries. (This can also be supported by other evidence)

In []: