

Life Satisfaction of People Living in Nature Parks

Context

This data science project contributes to the research project “Sustainable Quality of Life” currently conducted at the Centre for Development and Environment at the University of Bern. The overall objective of the research project “Sustainable Quality of Life” is to identify opportunities to link quality of life and sustainability. For this purpose, a written survey on life satisfaction and sustainability has been conducted in three nature parks (UNESCO Biospäre Entlebuch, Naturpark Gantrisch, Jurapark Aargau) and one control region.

Objectives

The objective of this data science project is to conduct a preliminary study and explore the data related to live satisfaction. More specifically it tries to answer the followin questions:

- Which groups show significant difference in life satisfaction?
- What factors impact people’s life satisfaction?

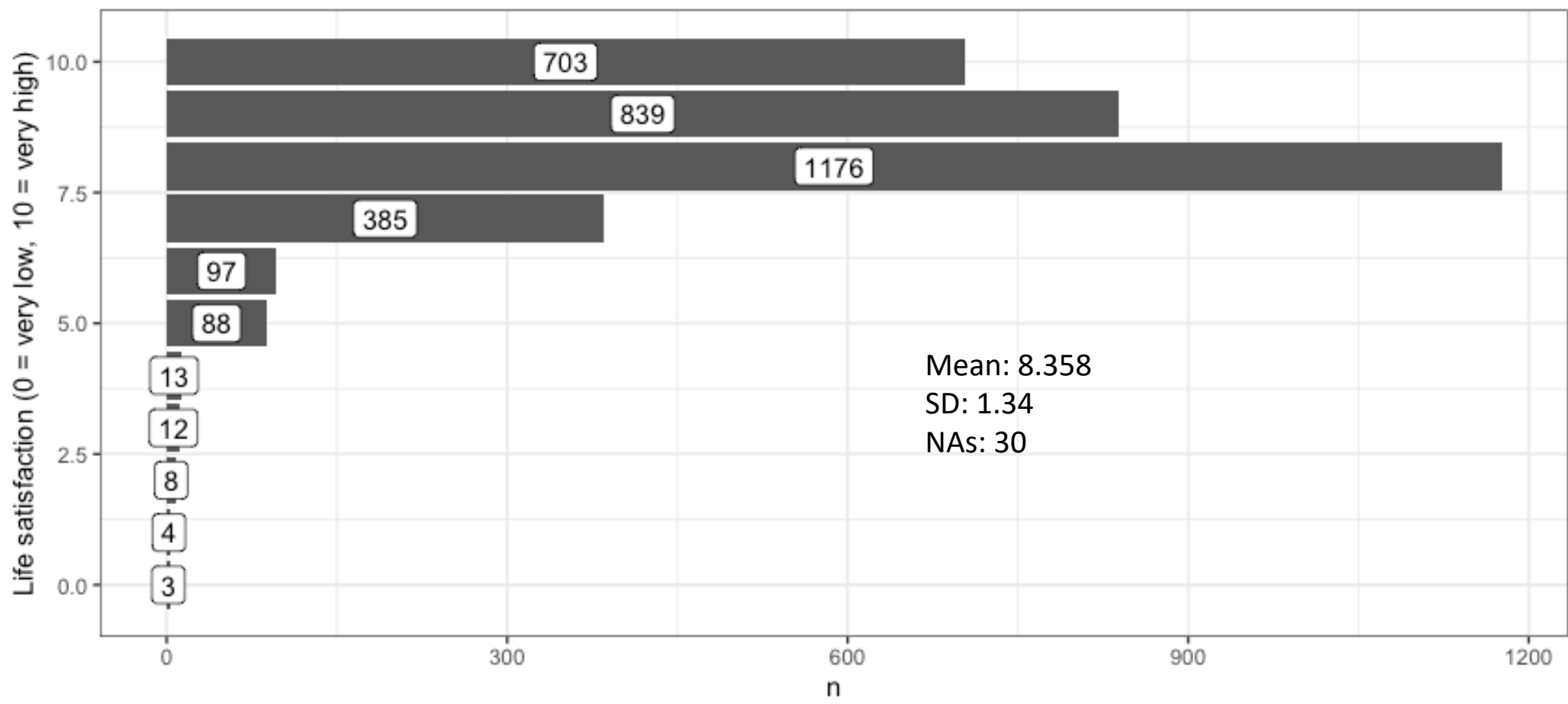
Data

Rate of return of valid questionnaires: 25% (out of 13’314)

Variable	Data type	Values	Missing values
Life satisfaction	Continuous	From 0 (= very) low to 10 (= very high)	0.89%
Region	Categorical	Park; Controll region	0.00%
Sex	Catecorical	Male (=0); Female (=1)	0.33%
Children	Categorical	Yes (=1) ; No (=0)	0.27%
Singel household	Categorical	Yes (=1) ; No (=0)	1.49%
Age	Continuous	From 18 to Inf	0.00%
Income	Continous	0 to 15000	18.15%

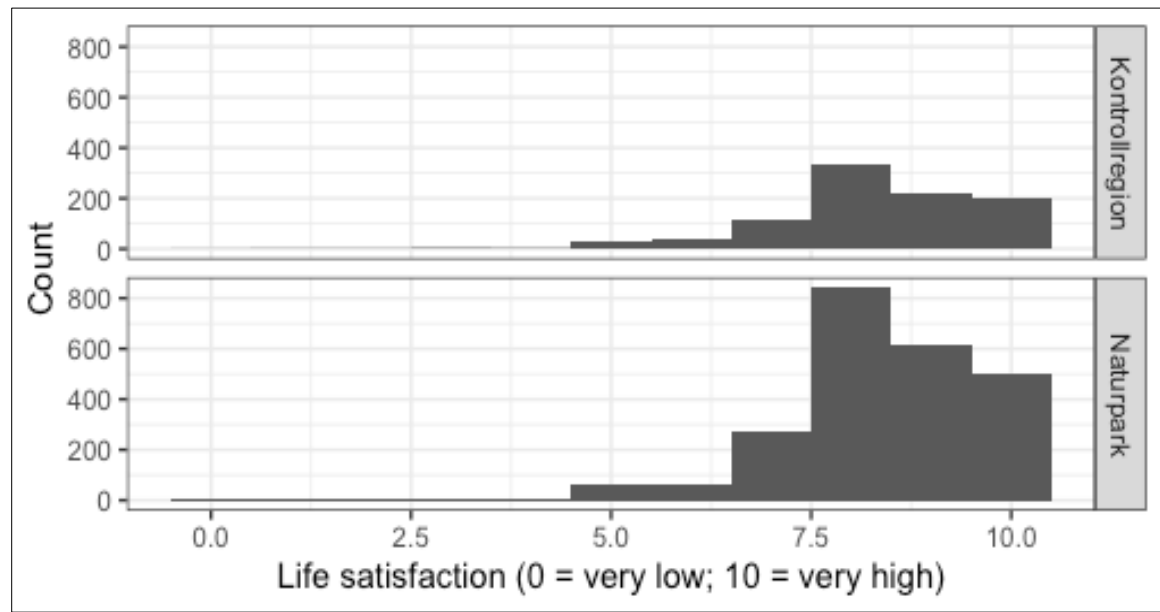
The data quality of most variables is considered to be sufficient. However, the income might have been transformed in a non-appropriate way and needs to be verified and possibly corrected.

Life satisfaction



Shapiro-Wilk normality test: W = 0.8598, p-value < 0.00000000000000022 The p-value is very small. We can reject the 0 hypothesis that the data has a normal distribution.

Life satisfaction by regions



Mann-Whitney U test
(2 groups, unpaired, non-parametric)
0 hypothesis = no difference

Life satisfaction by Region
Results: W = 1096739, p-value = 0.2735

There is no significant difference in life satisfaction between the study areas.

Life satisfaction by sex

Sex	Mean	SD
Male	8.305	1.356
Female	8.401	1.324

Mann-Whitney U test
(2 groups, unpaired, non-parametric)
0 hypothesis = no difference

Data: Life satisfaction by Sex
Results: W = 1314094, p-value = 0.03555

Women have a significantly higher life satisfaction then men.

Life satisfaction by children

Childern	Mean	SD
Yes	8.454	1.291
No	8.166	1.414

Mann-Whitney U test
(2 groups, unpaired, non-parametric)
0 hypothesis = no difference

Data: Life satisfaction by Children
Results: W = 1374798, p-value = 0.000000007068

Parents have a significantly higher life satisfaction then non-parents.

Life satisfaction by household

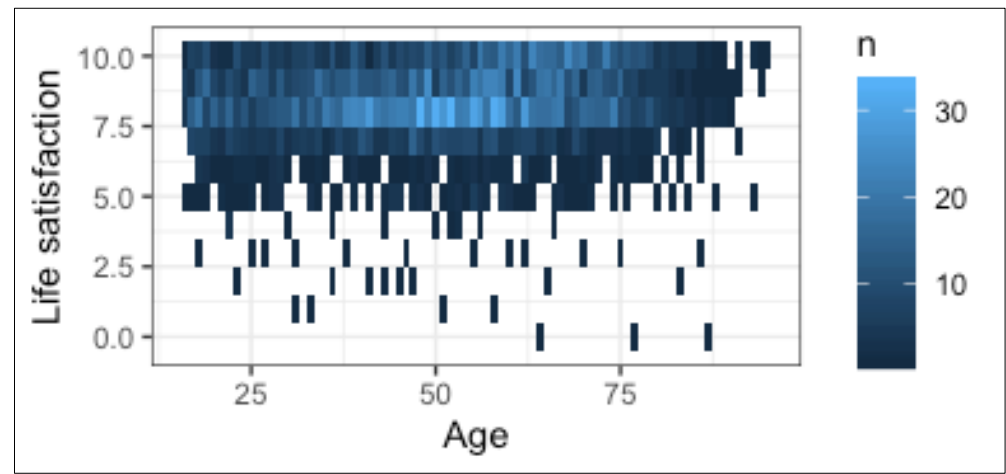
Singel hh	Mean	SD
Yes	8.079	1.711
No	8.398	1.268

Mann-Whitney U test
(2 groups, unpaired, non-parametric)
0 hypothesis = no difference

Data: Life satisfaction by Single household
Results: W = 561254, p-value = 0.002834

People in living in non-single households have a significantly higher life satisfaction people living in single households.

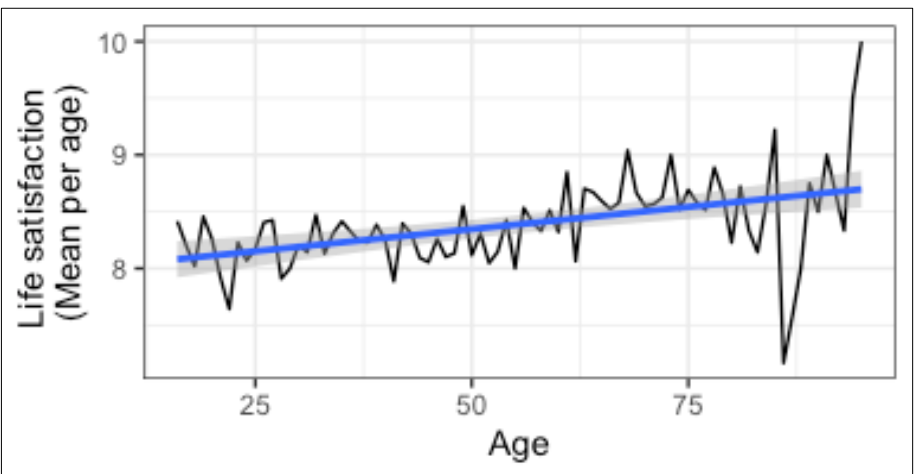
Life satisfaction by age



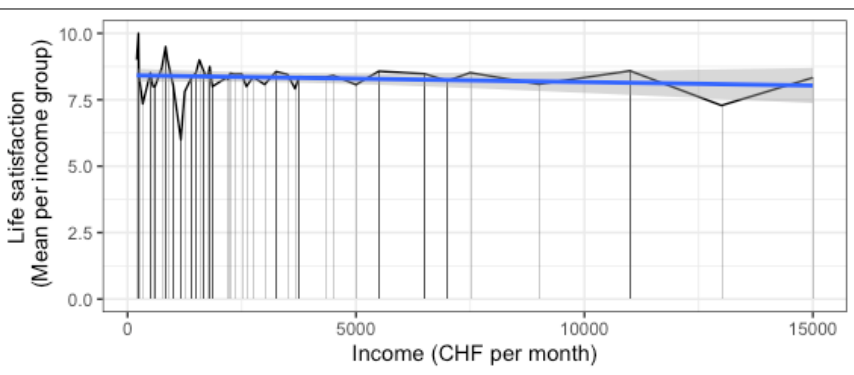
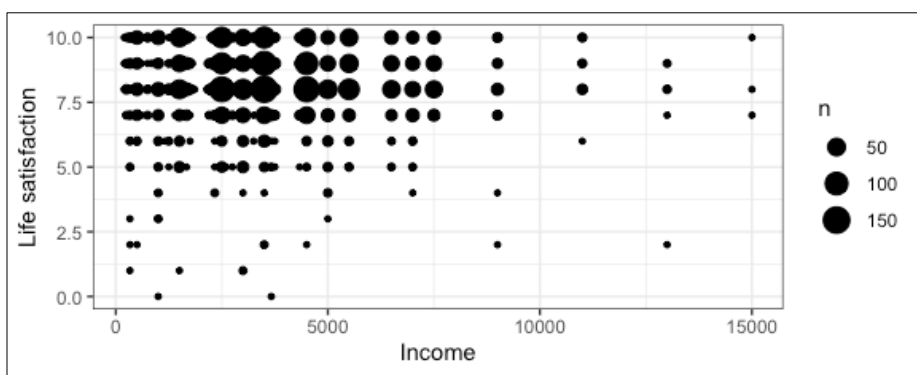
Kruskal-Wallis rank sum test
(3>= groups, unpaired, non-parametric)
0 hypothesis = no difference

Data: Life satisfaction by Age
Results: Kruskal-Wallis chi-squared = 176.93, df = 78, p-value = 0.000000001192

People with different ages have a significantly different life satisfaction.



Life satisfaction by income



Kruskal-Wallis rank sum test
(3>= groups, unpaired, non-parametric)
0 hypothesis = no difference

Data: Life satisfaction by Income
Kruskal-Wallis chi-squared = 56.849, df = 42, p-value = 0.06278
People with different incomes have not a significantly different life satisfaction.

First attempt of linear regression

Call:
lm(formula = Life satisfaction ~ Age + Sex + Children + Single hh + income, data = d)

Residuals:
Min 1Q Median 3Q Max
-8.5130 -0.5482 -0.0887 0.8346 2.4290

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.25547267 0.15907829 45.609 < 0.0000000000000002 ***
Age 0.01142662 0.00170323 6.709 0.0000000000238 ***
Sex 0.13212334 0.04933400 2.678 0.00745 **
Childden 0.03320145 0.06185513 -0.537 0.59148
Single hh 0.42078795 0.07660454 5.493 0.0000000431854 ***
Income 0.00002874 0.00001241 2.316 0.02062 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1

Residual standard error:
1.277 on 2713 degrees of freedom (639 observations deleted due to missingness)
Multiple R-squared: 0.03188, Adjusted R-squared: 0.03009
F-statistic: 17.87 on 5 and 2713 DF, p-value: < 0.00000000000000022

- The age, gender and household composition seem to have an significant impact on life satisfaction.
- However, the overall fit of the model is very low (R squared = 0.03009). I.e. the model explains very little of the variability
- Possible improvement: check better for assumptions, e.g. multicollinearity, normal distribution of error terms and homoscedasticity

Conclusion

- Life satisfaction is different according to different groups. For instance, significant differences among the population can be found according to the sex (female higher), whether people have children or not (higher with children), whether people live in a single household (higher if not) and according to their age (higher if older).
- The regression result do not (yet) provide good results and needs substantial improvement.
- Further interesting analysis include the inclusion of additional variables and the statistical inference to the entire population.