M05 – Mastery Project

Research Report

Darrell Gerber

December 16, 2021

# Summary

The Mastery project asks us to learn more about the quality of life in the US. The two main questions are 1) What factors impact the quality of life, and 2) Could the factors of quality of life vary from area to area. Dev10 provided two required datasets to use other data sets to support our hypothesis: a dataset of 8 Health-Related Quality of Life (HRQOL) measurements for all states and across 15 demographic categories (plus an overall measurement) another with population by zip code.

I narrowed down the 8 HRQOL measurements to three factors based on approaches used by other researchers: Mean Mentally Unhealthy Days, Mean Physically Unhealthy Days, and Percentage with Fair or Poor Self-Reported Health. These three HRQOLs are highly correlated, where states with a good rating in one measurement usually have an excellent rating in all measurements. However, the population relates poorly to the HRQOL measurements.

# Data Sources

## Required Datasets:

**Behavioral Risk Factor Data: Health-Related Quality of Life[[1]](#footnote-1)**

Summary - 1993 - 2010. Centers for Disease Control and Prevention (CDC). Data are from the Behavioral Risk Factor Surveillance System (BRFSS). All respondents to the BRFSS are non-institutionalized adults 18 years old or older. HRQOL surveillance helps identify unmet population health needs, including recognizing trends, disparities, and determinants of health in the population. HRQOL surveillance data is useful to inform decision-making and program and policy development. For example, program evaluation ensures that the population benefits from public health programs. A compact set of HRQOL measures, including a summary measure of unhealthy days, have been developed and validated for population health surveillance and have been widely used since 1993.

**US Population by Zip Code[[2]](#footnote-2)**

Summary: The United States census count (also known as the Decennial Census of Population and Housing) is a count of every resident of the US. Census data is publicly available through the census website, but much of the data is available in summarized data and graphs. Unfortunately, the raw data is often difficult to obtain, is typically divided by region, and must be processed and combined to provide information about the nation.

## Additional Datasets:

**Zip Code by State[[3]](#footnote-3)**

Summary: The IRS creates Individual Income Tax ZIP Code Data which contains selected income and tax items classified by State, ZIP Code, and size of adjusted gross income. This dataset connects zip codes to states.

# Quality of Life

According to the World Health Organization (WHO), Quality of life (QOL) is a broad multidimensional concept that usually includes subjective evaluations of positive and negative aspects of life. Health-Related Quality of Life (HRQOL) are dimensions related to public health.[[4]](#footnote-4)

The University of Wisconsin Population Health Institute and the Robert Wood Johnson Foundation publish the County Health Rankings (the Rankings) annually "to...demonstrate differences in health by place, raise awareness of the many factors that influence health, and stimulate community health improvement efforts."[[5]](#footnote-5) The Rankings use 34 measurements to create a composite score to rank each state's best and worst counties. The measurements relate to policies and programs, health factors, and health outcomes.

One of the categories in health outcomes is quality of life. Within that category, the Rankings use three measurements related to the HRQOL measurements included in our project data set.

|  |  |  |
| --- | --- | --- |
| Health Outcomes |  | |
| Quality of Life | **Measure** | **Data Source** |
| Poor or fair health | Behavioral Risk Factor Surveillance System |
| Poor physical health days | Behavioral Risk Factor Surveillance System |
| Poor mental health days | Behavioral Risk Factor Surveillance System |
| Low birthweight | National Center for Health Statistics |

Poor fair health, poor physical health days, and poor mental health days are widely used self-reported health statuses that help characterize the burden of disabilities and chronic diseases in a population. In addition to measuring how long people live, it is also essential to include measures that consider how healthy people are while alive. [[6]](#footnote-6) Mean Mentally Unhealthy Days, Mean Physically Unhealthy Days, and Percentage with Fair or Poor Self-Reported Health from our database relate closely to the measurements used by the Rankings and will be the targeted HRQOL measurements used throughout this report.

Low birth weight is an optional measurement for future revisions. The National Center for Health Statistics provides digital records of all births since 1968, including birthweight. Although, calculation of low birth weight (% of births below 2500g) is a secondary step after compiling the records.

## Could the factors of quality of life vary from area to area?

The Rankings find each year a significant variation from county to county.[[7]](#footnote-7)

|  |  |  |  |
| --- | --- | --- | --- |
| Measure | Least Healthy[[8]](#endnote-1) | Healthy[[9]](#endnote-2) | Ratio |
| Poor or fair health | 20% | 11% | 1.8 |
| Poor physical health days | 4.4 | 2.9 | 1.5 |
| Poor mental health days | 4 | 2.9 | 1.4 |

Based on their finding, we can expect similar variation from state to state in the dataset provided.

## What factors affect the quality of life?

We expect the overall HRQOL to vary by state, and the population varies state by state, so one question is, are the three HRQOL measurements and population are correlated. In other words, do high population states collectively have higher (or lower) health-related quality of life than low population states.

Via a simple thought experiment, we can expect the HRQOLs not to be related to the population. For example, Mississippi, Connecticut, and Iowa all have similar populations. Yet, Mississippi is notorious for its low quality of life, and Iowa and Connecticut are both known for having healthy people. So either Mississippi is an outlier, or the HRQOLs are poorly related to population.

# Analysis

A correlation matrix shows the relationship between the overall HRQOL measurements by state themselves and the population. (Table 1).

Table 1: Correlation matrix for HRQOLs and Population

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Column1*** | ***Mean mentally unhealthy days*** | ***Mean physically unhealthy days*** | ***Percentage with fair or poor self-rated health*** | ***Population*** |
| Mean mentally unhealthy days | 1 |  |  |  |
| Mean physically unhealthy days | 0.847226657 | 1 |  |  |
| Percentage with fair or poor self-rated health | 0.735478542 | 0.912840515 | 1 |  |
| Population | 0.209809028 | 0.150785978 | 0.204486506 | 1 |

The correlation between the HRQOLs and each other is high, while the correlation of the HRQOLs with population is weak. The scatterplots in Table 2 support the calculated correlations. The scatterplots of HRQOLs to each other (in orange) tightly cluster around the linear trendline. On the other hand, the scatterplots of each HRQOL to population vary widely, illustrating a high residual error from the trendline.

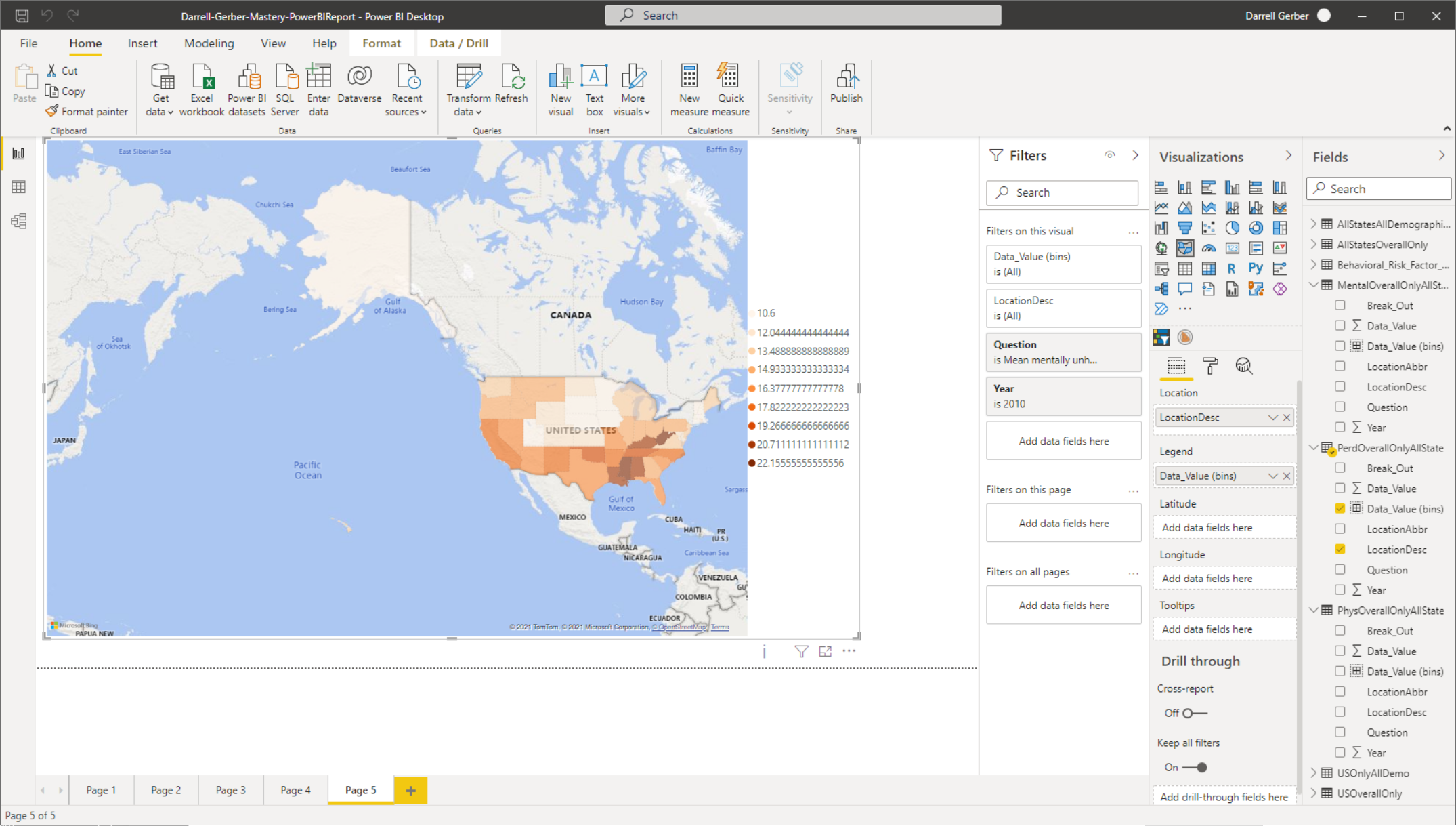
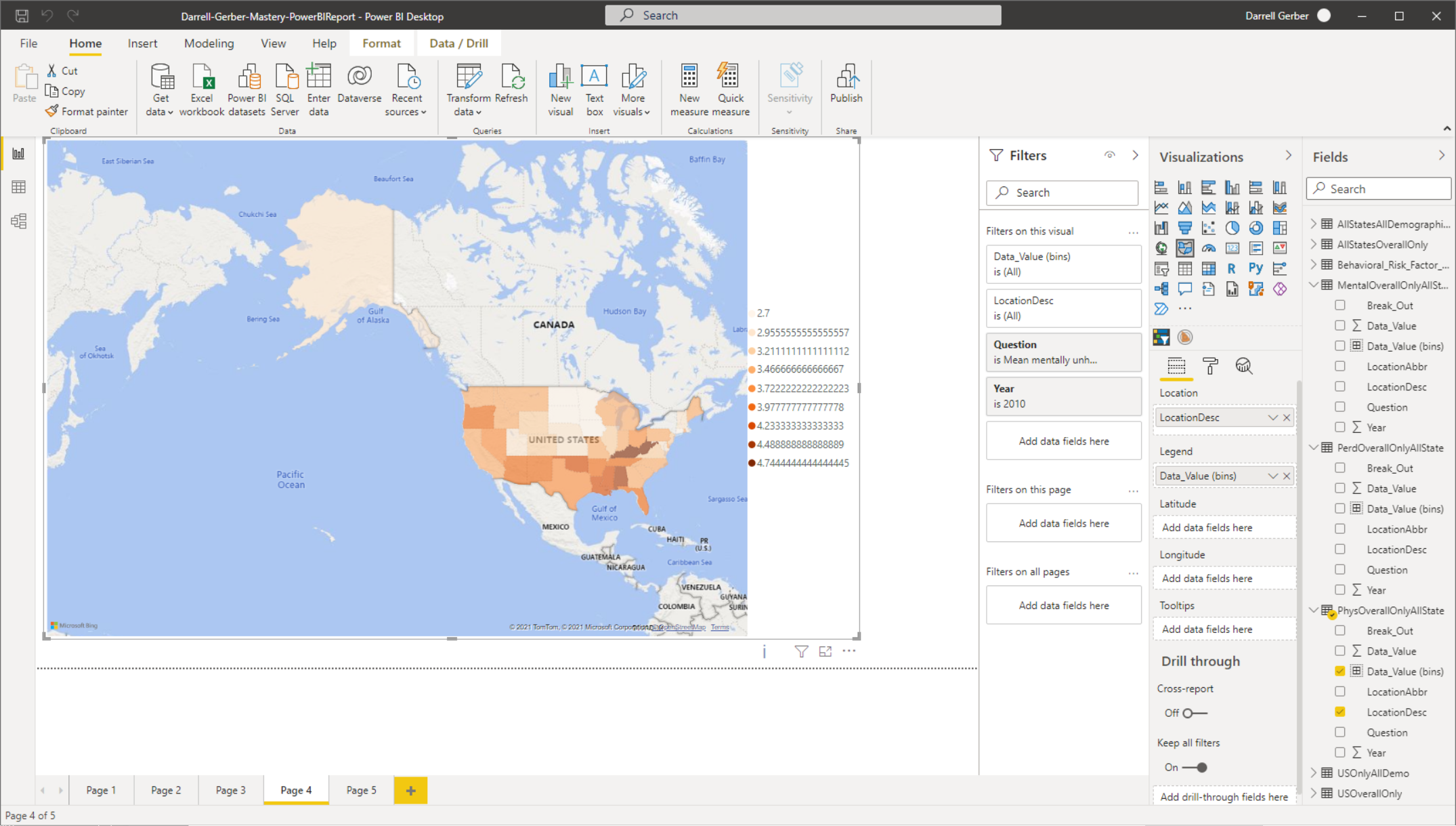
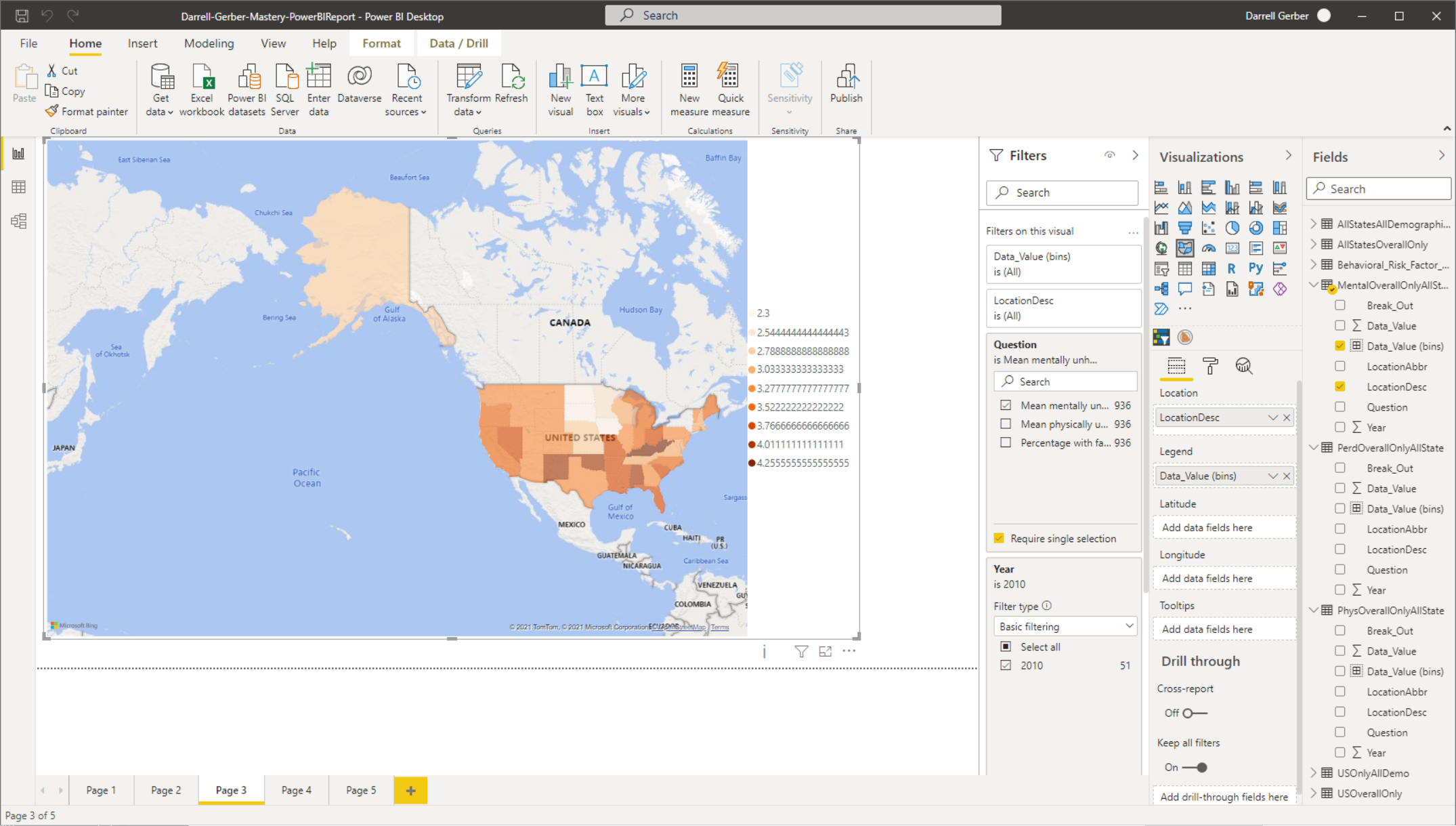
Table 2: Scatterplots illustrating the correlation between HRQOLs and population.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Mean Mentally Unhealthy Days* | *Mean Physically Unhealthy Days* | *Percentage with Fair or Poor Self-Rated Health* |
| *Mean Physically Unhealthy Days* |  |  |  |
| *Percentage with Fair or Poor Self-Rated Health* |  |  |  |
| *Population (x Millions)* |  |  |  |

As expected, the three HRQOL measurements correlate highly with each other, meaning that if a state has a high value in one, they very likely have a high value in the other HRQOLs. Again, it is unsurprising that HRQOLs are highly related because of the interconnected causal factors that influence mental and physical health.

Also expected is the lack of correlation between the population and the HRQOLs. The scatterplots show that the most populous states appear to be outliers. However, eliminating them would not likely affect the trendlines since the high population states lie near the mean of each HRQOL.

Another hypothesis is that the HRQOLs vary by state. Figure 1 illustrates that the HRQOL measurements do indeed vary by state.



A. B. C.

Figure 1: Health-Related Quality of Life measurements by the state for 2010. A. Mean Mentally Unhealthy Day B. Mean Physically Unhealthy Days, and C. Percentage with Fair or Poor Self-Reported Health.[[10]](#footnote-8)

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Mean mentally unhealthy days* | *Mean physically unhealthy days* | *Percentage with fair or poor self-rated health* |
| Units | days | days | percent of surveyed |
| Mean | 3.429411765 | 3.590196078 | 15.34313725 |
| Standard Error | 0.068958306 | 0.076350836 | 0.470004213 |
| Median | 3.5 | 3.6 | 14.7 |
| Mode | 3.5 | 3.7 | 11.5 |
| Standard Deviation | 0.492460808 | 0.545254033 | 3.356501447 |
| Sample Variance | 0.242517647 | 0.297301961 | 11.26610196 |
| Kurtosis | -0.009278314 | -0.085604258 | -0.132789531 |
| Skewness | -0.306453339 | 0.447679089 | 0.697542672 |
| Range | 2.2 | 2.3 | 13 |
| Minimum | 2.3 | 2.7 | 10.6 |
| Maximum | 4.5 | 5 | 23.6 |
| Sum | 174.9 | 183.1 | 782.5 |
| Count | 51 | 51 | 51 |
| Confidence Level(95.0%) | 0.138506834 | 0.153355168 | 0.944031244 |

Descriptive statistical analysis reveals that, while the values for the HRQOLs vary by state, they remain clustered around a central value – the confidence level for all is small compared to the mean. Additionally, the HRQOLs are skewed, making it more difficult to differentiate between states one the extremes.

# Conclusion

The provided CDC dataset contains 8 Health-Related Quality of Life measurements. Of those, three were selected as the most representative when comparing the quality of life by state: Mean mentally unhealthy days, Mean physically unhealthy days, and Percentage with fair or poor self-rated health. It was hypothesized and confirmed that the HRQOLs are highly related but poorly associated with the state population.

The HRQOLs vary from state to state, making them helpful in evaluating and ranking the quality of life in states. Potential applications include exploring the variations by the state to make decisions, including relocation, business placement, public policy changes, funding allocations, and assessing community vulnerability.

1. Centers for Disease Control. Behavioral Risk Factor HRQOL (1993-2010) [Data file]. Retrieved from <https://data.world/cdc/behavioral-risk-factor-hrqol>. Accessed December 13, 2021. [↑](#footnote-ref-1)
2. US Census Bureau. (2000 and 2010). US Population By Zip Code (Version 1) [Data file]. Retrieved from <https://www.kaggle.com/census/us-population-by-zip-code>. Accessed December 13, 2021. [↑](#footnote-ref-2)
3. National Bureau of Economic Research. Individual Income Tax Statistics - ZIP Code Data (SOI) (2017-06-21) [Data file]. Retrieved years 2005 and 2010 from https://www.kaggle.com/census/us-population-by-zip-code [↑](#footnote-ref-3)
4. The WHOQOL Group. The World Health Organization Quality of Life Assessment (WHOQOL). Development and psychometric properties. Soc Sci Med 1998;46:1569-1585. [↑](#footnote-ref-4)
5. Remington, P.L., Catlin, B.B. & Gennuso, K.P. The County Health Rankings: rationale and methods. Popul Health Metrics 13, 11 (2015). https://doi.org/10.1186/s12963-015-0044-2 [↑](#footnote-ref-5)
6. County Health Rankings and Roadmaps, <https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-outcomes/quality-of-life/poor-or-fair-health>, Accessed December 16, 2021. [↑](#footnote-ref-6)
7. Remington, pg. 8. [↑](#footnote-ref-7)
8. The average of the 5 least healthy counties from each state [↑](#endnote-ref-1)
9. The average of the 5 most healthy counties from each state [↑](#endnote-ref-2)
10. Using ColorBrewer Oranges color scheme for 9 bins. <https://colorbrewer2.org/#type=sequential&scheme=Oranges&n=9> [↑](#footnote-ref-8)