

# Introduction to Health Data Science Report

Melanie Rees-Roberts

2nd December 2024

## Understanding health and wealth in Scottish Council areas.

### Background:

Over the last 100 years, life expectancy in the UK has increased significantly with advances in living standards and advancing healthcare (1). However, over the last 10 years this has stalled and even declined in some instances alongside significant inequalities arising across the country (2-3). In Scotland, disparities in life expectancy are evident where the most deprived exhibit approximately 10 years difference in life expectancy compared to their more affluent neighbourhoods (4). In recent years, the cost of living crisis has also impacted on peoples health and wellbeing but the extent to which this has impacted on different income levels is unknown and thereafter how this links to health (5-6). Available data in the Scottish Public Health Observatory may held to explore the extent to which people feel able to manage their own money and how it may act as a useful proxy measure for financial stress within a cost of living crisis regardless of income levels.

### Research questions:

1. What is the variation in health and mental health across different Scottish council areas?
2. How is population income associated with the perception of being able to manage money well?
3. How does the ability to manage money well impact on health and mental health in Scotland?

### Packages and libraries:

The following packages were used:

- `install.packages("tidyverse")`
- `install.packages("car")`

The following libraries were used:

- `library(tidyverse)`
- `library(readr)`
- `library(tidyr)`
- `library(dplyr)`

- library(ggplot2)
- library(stats)
- library(car)

## Data acquisition

In this work, we used the datasets listed below. Full descriptions of each dataset can be found on the Indicator definitions and schedule page of the Scottish Public Health Observatory website.

- **Population income deprived** - Number and percentage of total population classified as income deprived within SIMD income domain. Calculated based on 2011 estimates. Last updated August 2021.
- **Healthy life expectancy, females** - Healthy life expectancy (HLE) at birth for females. Last updated June 2023.
- **Healthy life expectancy, males** - Healthy life expectancy (HLE) at birth for males. Last updated June 2023.
- **Households managing well financially** - Percentage of households managing very or quite well financially these days. Respondents were asked “Taking everything together, which of these phrases on this card best describes how you and your household are managing financially these days?” and the six responses ranged from “manage very well” to “are in deep financial trouble”. The indicator is the percentage who responded either “manage very well” or “manage quite well”. Last updated September 2024.
- **Mental wellbeing score** - Mean score on the WEMWBS scale (adults, 16+ years). WEMWBS stands for Warwick-Edinburgh Mental Wellbeing Scale. The questionnaire consists of 14 positively worded items designed to assess: positive affect (optimism, cheerfulness, relaxation) and satisfying interpersonal relationships and positive functioning (energy, clear thinking, self-acceptance, personal development, mastery and autonomy). It is scored by summing the response to each item answered on a 1 to 5 Likert scale (‘none of the time’, ‘rarely’, ‘some of the time’, ‘often’, ‘all of the time’). The total score ranges from 14 to 70 with higher scores indicating greater wellbeing. The SIMD figures were age-standardised to the population of Scotland to aid comparison between the quintiles. Last updated December 2023

These data were downloaded from [https://scotland.shinyapps.io/ScotPHO\\_profiles\\_tool/](https://scotland.shinyapps.io/ScotPHO_profiles_tool/) and imported into R Studio in a single file.

```
Data <- read_csv("ScotPHO_data.csv")
glimpse(Data)
```

## Data cleaning and preparation

Data was manipulated to create tidy datasets as each variable was listed in a single column with the name of the indicator in the column ‘indicator’. For each of the four indicators used, tidy datasets were created.

### Dataset 1 - Healthy life expectancy.

Create tidy data with healthy life expectancy for males and females.

```

hle <- Data %>%
  filter(grepl('Healthy', indicator)) %>%
  select(area_code, area_name, indicator, measure, upper_confidence_interval,
         lower_confidence_interval) %>%
  rename(sex = indicator,
         hle = measure) %>%
  mutate(sex = gsub('Healthy life expectancy, females', '0', sex,)) %>%
  mutate(sex = gsub('Healthy life expectancy, males', '1', sex,))

head(hle)

```

```

## # A tibble: 6 x 6
##   area_code area_name sex    hle upper_confidence_int~1 lower_confidence_int~2
##   <chr>      <chr>   <chr> <dbl>          <dbl>          <dbl>
## 1 S12000005 Clackmann~ 0      59.6          63.3          55.9
## 2 S12000006 Dumfries ~ 0      59.2          61.9          56.6
## 3 S12000008 East Ayrs~ 0      59.9          62.4          57.4
## 4 S12000010 East Loth~ 0      68.4          71           65.8
## 5 S12000011 East Renf~ 0      67.2          69.9          64.4
## 6 S12000013 Na h-Eile~ 0      64           68.2          59.8
## # i abbreviated names: 1: upper_confidence_interval,
## # 2: lower_confidence_interval

```

The data was mutated to create a single mean healthy life expectancy for males and females combined for each Council area.

```

mean_hle <- hle %>%
  group_by(area_code) %>%
  mutate(mean_hle = mean(hle))

```

This led to duplicate mean healthy life expectancy values, so the data was sliced to remove the rows containing sex = 0 and remove the column sex.

```

mean_hle1 <- mean_hle %>%
  filter(sex != 1) %>%
  select(area_code, area_name, mean_hle)

head(mean_hle1)

```

```

## # A tibble: 6 x 3
## # Groups:   area_code [6]
##   area_code area_name      mean_hle
##   <chr>      <chr>          <dbl>
## 1 S12000005 Clackmannanshire  60.4
## 2 S12000006 Dumfries & Galloway 60.8
## 3 S12000008 East Ayrshire      58.4
## 4 S12000010 East Lothian      64.6
## 5 S12000011 East Renfrewshire 67.0
## 6 S12000013 Na h-Eileanan Siar 65.7

```

## Dataset 2 - Population income deprived

The process was repeated for the remaining datasets.

```
pid <- Data %>%
  drop_na()

pid <- pid %>%
  select(area_code, numerator, measure) %>%
  rename(population = numerator,
         pid = measure)
```

### Dataset 3 - Mental health score deprived

```
mental <- Data %>%
  filter(grepl('Mental', indicator)) %>%
  select(area_code, measure) %>%
  rename(mental_score = measure)
```

### Dataset 4 - Households managing financially well

```
finance <- Data %>%
  filter(grepl('Households', indicator)) %>%
  select(area_code, measure) %>%
  rename(finance = measure)
```

All indicators were then combined into a single table.

```
data_final <- inner_join(mean_hle1, pid, by = 'area_code')

data_final <- inner_join(data_final, mental, by = 'area_code')

data_final <- inner_join(data_final, finance, by = 'area_code')

head(data_final)
```

```
## # A tibble: 6 x 7
## # Groups:   area_code [6]
##   area_code area_name      mean_hle population   pid mental_score finance
##   <chr>      <chr>          <dbl>      <dbl> <dbl>      <dbl>    <dbl>
## 1 S12000005 Clackmannanshire    60.4        7305  14.2        46.5    45.5
## 2 S12000006 Dumfries & Galloway  60.8       17084  11.5        48.3    51.6
## 3 S12000008 East Ayrshire      58.4       18532  15.2        45.7    43.5
## 4 S12000010 East Lothian      64.6       10711  10.2         47     49.2
## 5 S12000011 East Renfrewshire  67.0        6424   6.8        48.7     60
## 6 S12000013 Na h-Eileanan Siar  65.7        2637   9.8        50.3    67.5
```

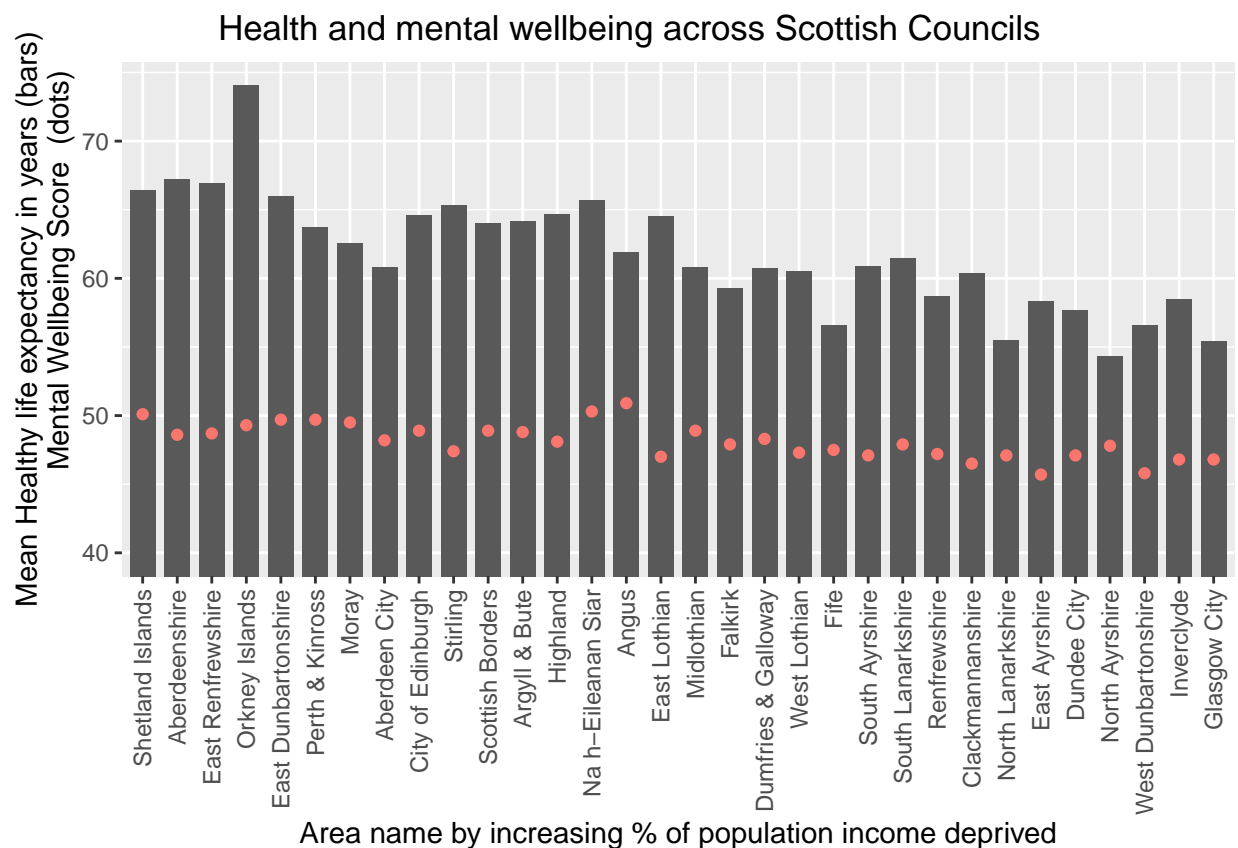
## Data analysis and visualisation

### 1. What is the variation in health and mental health across different Scottish council areas?

To visualise the differences in health and mental health across Scottish councils we plotted each council area in a column plot with councils ordered on the x axis in order of increasing % of the population that were income deprived (i.e. income deprivation increases with the x axis). On top of this, mental wellbeing scores were plotted using points showing the mental health scores of these council areas.

Figure 1: The variation in health and mental health score across different Scottish council areas.

```
data_final %>%
  ggplot() +
  geom_col(aes(x = reorder(area_name, pid), y = mean_hle),
           width=0.75, position=position_dodge(0.5)) +
  geom_point(aes(x=area_name, y = mental_score, colour = "red")) +
  coord_cartesian(ylim = c(40, NA)) +
  ggtitle("Health and mental wellbeing across Scottish Councils") +
  labs(x="Area name by increasing % of population income deprived",
       y = "Mean Healthy life expectancy in years (bars)
           Mental Wellbeing Score (dots)") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
        legend.position = "none")
```



From these data you can see that there is a general decrease in mean healthy life expectancy with increasing

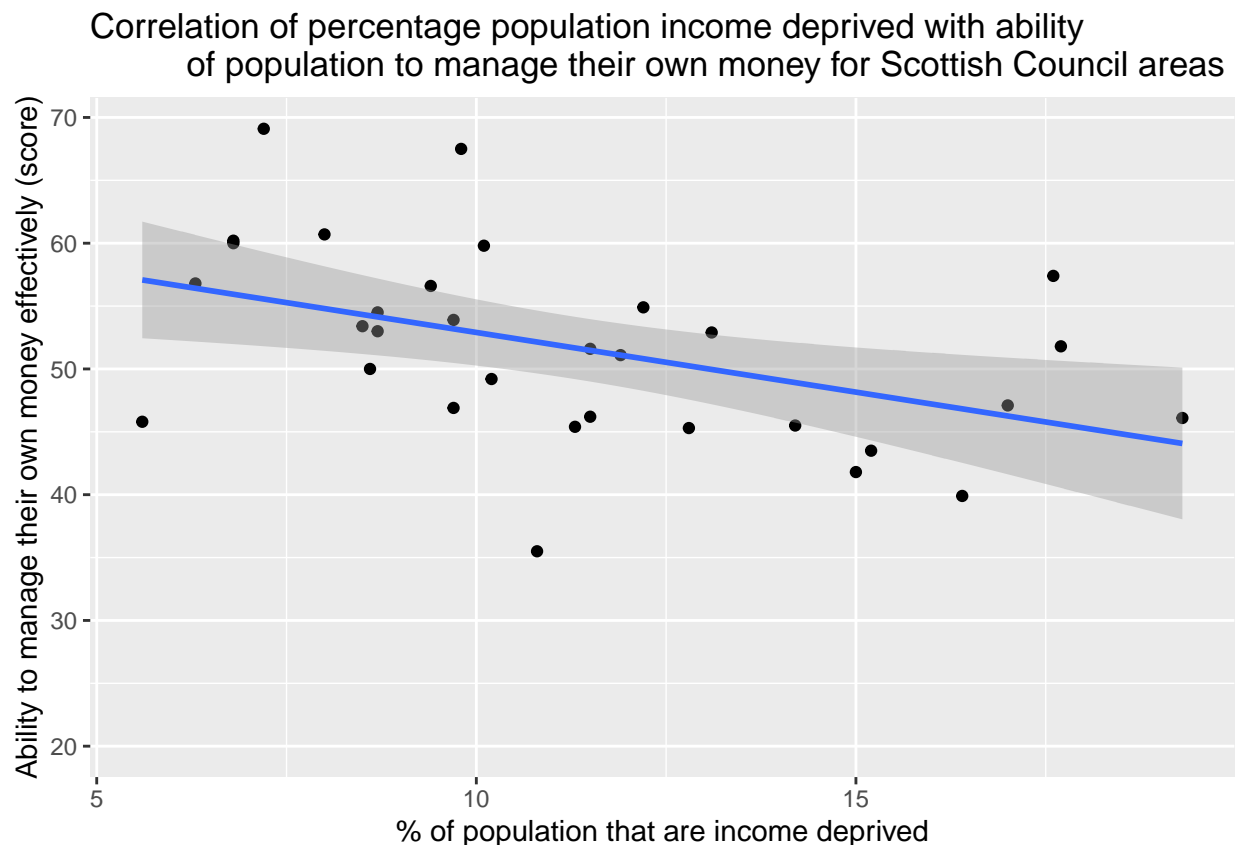
% of the population that is income deprived. The Orkney Islands stand out as being an outlier to this trend with a life expectancy of approximately 10 years higher than its comparable council areas. Mental wellbeing score, however, does not exhibit a similar correlation with income deprivation and is generally consistent across council areas with some variation in individual cases e.g. Angus having the highest mental wellbeing score and East Lothian one of the lowest despite similar income deprivation levels.

## 2. How is population income associated with the perception of being able to manage money well?

To understand if income deprivation correlates with the ability to manage money, we compared % of the population that is deprived with the ability of the population to manage their money well variable as a scatter plot with best fit line added.

Figure 2: Correlating the % of the population that are income deprived with the ability to manage money well in Scottish council areas

```
data_final %>%
  ggplot(aes(x = pid, y = finance)) +
  geom_point() +
  geom_smooth(method = lm) +
  ggtitle("Correlation of percentage population income deprived with ability
    of population to manage their own money for Scottish Council areas") +
  labs(x="% of population that are income deprived",
    y = "Ability to manage their own money effectively (score)") +
  coord_cartesian(ylim = c(20, NA))
```



The graph shows a negative correlation indicating that those council areas where there is a higher proportion of people income deprived also manage their money less well.

### 3. How does the ability to manage money well impact on health and mental health in Scotland?

#### Predicting mean healthy life expectancy using the % of the population income deprived and ability to manage finances well

First, the data were used to create a multiple linear regression model of mean healthy life expectancy predicted by population income deprivation and the ability to manage money well.

```
q3_Model = lm(formula = mean_hle ~ pid + finance, data = data_final)
summary(q3_Model)
```

```
##
## Call:
## lm(formula = mean_hle ~ pid + finance, data = data_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6741 -1.6949  0.1929  1.2022  7.3068
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  66.45058    3.87392   17.153 < 2e-16 ***
## pid         -0.89344    0.12154   -7.351 4.25e-08 ***
## finance      0.10578    0.05859    1.805  0.0814 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.203 on 29 degrees of freedom
## Multiple R-squared:  0.7518, Adjusted R-squared:  0.7347
## F-statistic: 43.92 on 2 and 29 DF,  p-value: 1.68e-09
```

It is important to check for the normality of the residuals to understand the model fit and explore any multi-collinearity. In this model, our residuals were normal in distribution and the collinearity for variables pid and finance were below 2.5 indicating no considerable collinearity present in the model between independent variables.

```
q3_Model_residuals = q3_Model$residuals # Get the model residuals
hist(q3_Model_residuals) # Plot the result
qqnorm(q3_Model_residuals) # Plot the residuals
qqline(q3_Model_residuals) # Plot the Q-Q line

# look at collinearity
vif_Values <- vif(q3_Model)
vif_Values %>%
  barplot(main = "VIF Values", horiz = TRUE, col = "steelblue") #create horizontal
#bar chart to display each VIF value
```

The equation for our model can be found in the summary. For this regression it is:

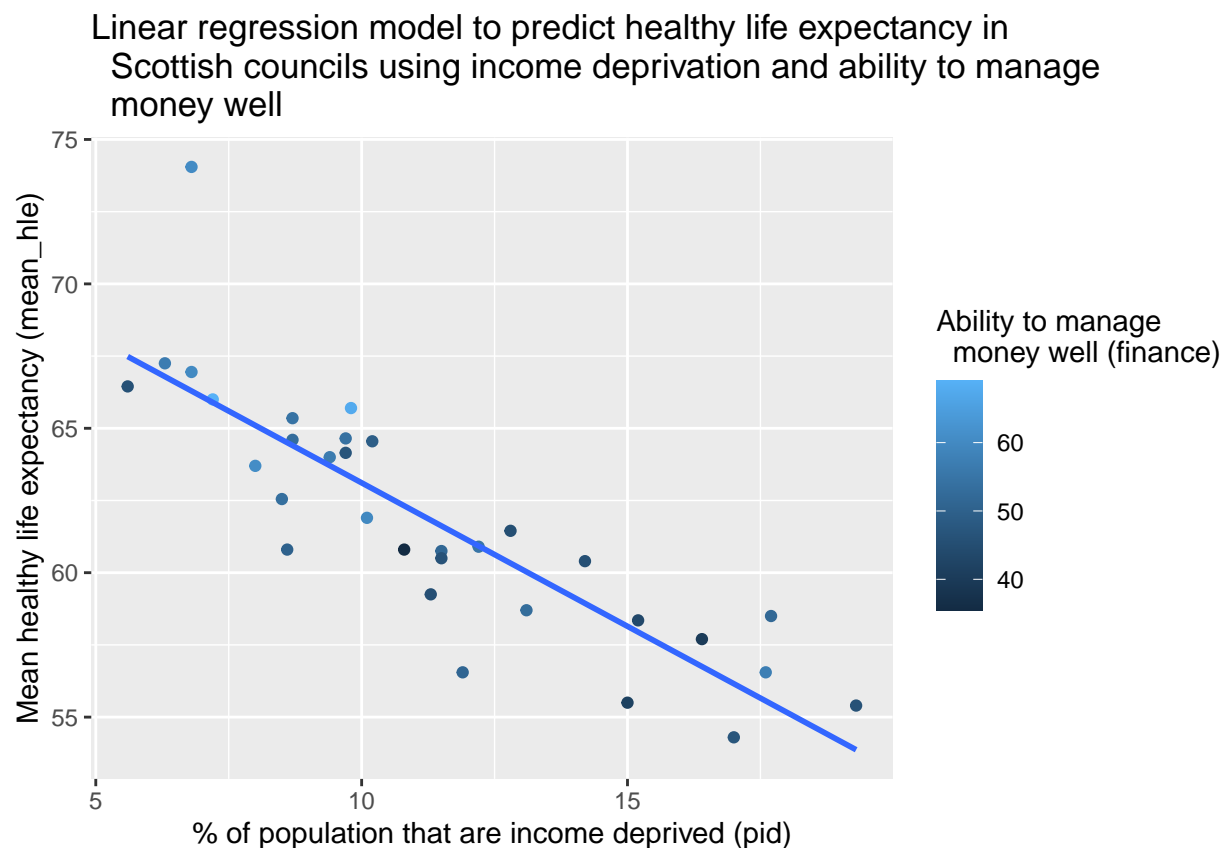
mean healthy life expectancy =  $66.5 - 0.89x + 0.11y$

where x is our population income deprived and y is the ability to manage own finance variable.

The adjusted  $R$ -squared value is  $r=0.7347$ . This indicates a moderate to borderline strong correlation between the variables. Figure 3 below visualised the relationship.

Figure 3: Predicting healthy life expectancy in Scottish council areas using income deprivation and financial management ability

```
data_final %>%
ggplot(aes(y=mean_hle,x=pid,colour=finance)) +
  geom_point() +
  stat_smooth(method="lm", se=FALSE) +
  ggtitle("Linear regression model to predict healthy life expectancy in
Scottish councils using income deprivation and ability to manage
money well") +
  labs(x="% of population that are income deprived (pid)",
       y = "Mean healthy life expectancy (mean_hle)",
       colour = "Ability to manage
money well (finance)")
```



**Predicting mental wellbeing score in Scottish council areas using the % of the population income deprived and ability to manage finances well**

The modelling was repeated to predict mental wellbeing score using the % of the population income deprived and the ability to manage finances well.



```
q3_Model_mental = lm(formula = mental_score ~ pid + finance, data = data_final)
summary(q3_Model_mental)
```

```
##
## Call:
## lm(formula = mental_score ~ pid + finance, data = data_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4049 -0.6287  0.1159  0.5273  2.2489
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  49.07956    1.55065   31.651 < 2e-16 ***
## pid         -0.22950    0.04865   -4.717 5.56e-05 ***
## finance      0.03160    0.02345    1.347  0.188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8817 on 29 degrees of freedom
## Multiple R-squared:  0.5659, Adjusted R-squared:  0.5359
## F-statistic: 18.9 on 2 and 29 DF,  p-value: 5.566e-06
```

```
q3_Model_mental_residuals = q3_Model$residuals # Get the model residuals
hist(q3_Model_mental_residuals) # Plot the result
qqnorm(q3_Model_mental_residuals) # Plot the residuals
qqline(q3_Model_mental_residuals) # Plot the Q-Q line

# look at collinearity
vif_Values2 <- vif(q3_Model_mental)
vif_Values2 %>%
  barplot(main = "VIF Values", horiz = TRUE, col = "steelblue") #create horizontal
#bar chart to display each VIF value
```

The residuals were normal and no concerns around collinearity were identified.

The equation for this model is:

Mental wellbeing score = 49.1 - 0.23x + 0.03y

where x is our population income deprived and y is the ability to manage own finance variable.

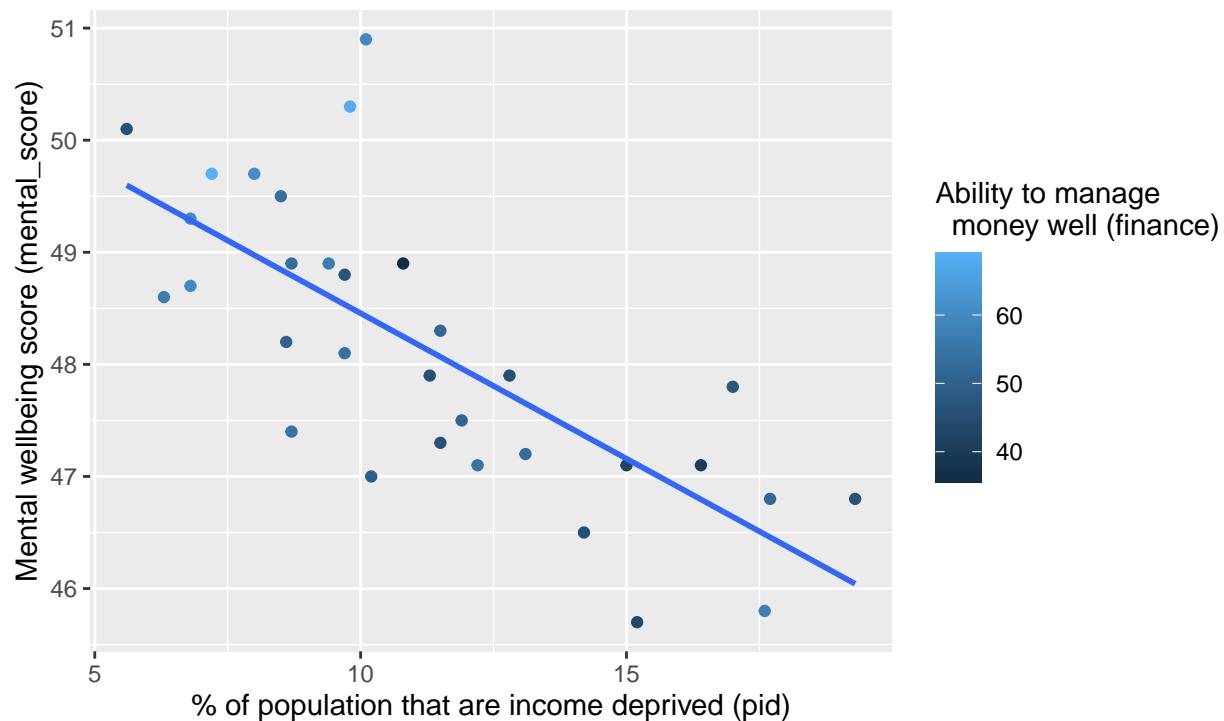
The adjusted *R*-squared value is  $r=0.5359$ . This indicates a weak relationship between the variables. Figure 4 below visualises the relationship. As demonstrated in the plot for research question 1, mental health seems to be less associated with income deprivation.

*Figure 4: Predicting mental health score using income deprivation and financial management ability*

```
data_final %>%
  ggplot(aes(y=mental_score,x=pid,colour=finance)) +
  geom_point() +
  stat_smooth(method="lm", se=FALSE) +
  ggtitle("Linear regression model to predict mental wellbeing score in
  Scottish councils using income deprivation and ability to manage
  money well") +
```

```
labs(x="% of population that are income deprived (pid)",
     y = "Mental wellbeing score (mental_score)",
     colour = "Ability to manage
money well (finance)")
```

Linear regression model to predict mental wellbeing score in Scottish councils using income deprivation and ability to manage money well



## Report summary

In this report we have explored the variation of healthy life expectancy and mental wellbeing across the 32 Scottish council areas. There is variation across Scottish council areas in healthy life expectancy but minimal or small amounts of variation in mental wellbeing scores. Across Scottish council areas, a trend emerged showing that those areas with the least proportion of its population being classified as income deprived (least deprived) had the highest healthy life expectancy. The most deprived areas had the lowest healthy life expectancy showing a considerable inequality in healthy life expectancy of 10 years. The Orkney Islands stood out as a council area that did not conform to this trend. Although it was difficult to visualise a similar trend for mental wellbeing score, some council areas with similar levels of deprivation, for example Angus and East Lothian, had disparate mental wellbeing of their populations. Understanding and observing trends in life expectancy and mental health alongside deprivation and other contextual factors may support policies to improve the general health of Scottish council areas. These results suggest a focus on reducing deprivation would improve healthy life expectancy rather than mental wellbeing.

Given the current cost of living crisis, we then went on to understand if a variable that measures households ability to manage money well would correlate with the same areas deprivation level. There was a negative correlation between the ability of households to manage their money well with increasing income deprivation. This shows that despite the context of a cost of living crisis, those areas with higher proportions of the

population being income deprived were still less able to manage their money well. However, we did see a few council areas that were outliers, for example where areas of least deprivation had higher and lower abilities in managing money well.

The data was then used to create models for predicting healthy life expectancy and mental wellbeing using the percentage of the population that were income deprived and the ability to manage money well. The models showed that these predictors were better at estimating healthy life expectancy ( $r = 0.75$ ) compared to mental wellbeing ( $r = 0.53$ ). These models could help to inform policy and targets for improving healthy life expectancy in council areas where deprivation and healthy life expectancy is low.

Limitations to this analysis included data variables collected and updated in different years. Therefore, the exact ability of the datasets to reflect each variable at one specific point in time is not possible and should be considered when interpreting this report. Furthermore, the variables used, although, similar in scale result from different methods of computation whereby mental wellbeing score and ability to manage money well are likert scores converted to a continuous scale using percentages. Therefore, given the subjectiveness of answering questions on a likert scale and differing scales, caution should be taken when interpreting these variables. Finally, we did not perform statistical testing to confirm correlations performed or predictor variables in the models created.

## References

1. Office for National Statistics. How has life expectancy changed over time?. 9 September 2015. Accessed 2.12.24.
2. The Health Foundation. Mortality and life expectancy trends in the UK. November 2019. Accessed 2.12.24.
3. Department for Health and Social Care, UK government. Chief Medical Officer's Report 2021: health in coastal communities. 21 July 2021. Accessed on 2.12.24.
4. National Records of Scotland. Life expectancy in Scotland 2020-2022, provisional figures. 26 September 2023. Accessed 2.12.24.
5. Institute for Government. Cost of living crisis. 7 February 2022. Accessed 2.12.24.
6. The Health Foundation. The cost-of-living crisis is a health emergency too. 7 October 2022. Accessed 2.12.24.