03_modeling_risk_factors

Tamara

2025-06-15

In this section, I use a series of linear models to understand how socioeconomic need (EQI) and volatility in attendance jointly predict average attendance. These models build upon one another, helping to answer: How much better can we predict attendance when we include volatility as a risk factor?

Step 1: Base Model – Only EQI

```
## [1] 0.6354139
```

This shows a strong negative association:

- **higher EQI scores (greater socioeconomic need) predict lower attendance**.
- The model explains about **63.5%** of the variance in attendance.

Step 2: Adding volatility to the model

```
## [1] 0.9436893
```

Volatility is a powerful predictor: students in more unstable attendance environments tend to have significantly lower average attendance.

- The model fit improves dramatically: **R^2 increases from 63.5% to 94.4%**.
- This shows that **volatility is not just noise — it explains an additional 30.9% of the variance** beyond EQI.

Step 3: Adding the interaction

```
## [1] 0.9535556
```

This interaction is significant and improves model fit again (R^2 = 95.4%). It suggests that the impact of EQI on attendance is amplified in more volatile regions. In other words:

> In regions with unstable attendance, socioeconomic disadvantage has an even stronger link to low attendance.

```
## Analysis of Variance Table
##
## Model 1: avg_present ~ eqi_mean
## Model 2: avg_present ~ eqi_mean + volatility_present
## Model 3: avg_present ~ eqi_mean * volatility_present
##   Res.Df     RSS Df Sum of Sq        F    Pr(>F)
## 1   2782 2702.82
## 2   2781  417.45  1   2285.37 18452.30 < 2.2e-16 ***
## 3   2780  344.31  1     73.14   590.57 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## # A tibble: 3 x 2
##   Model             R_squared
##   <chr>                 <dbl>
## 1 EQI only              0.635
## 2 EQI + Volatility      0.944
## 3 EQI + Interaction     0.954
```

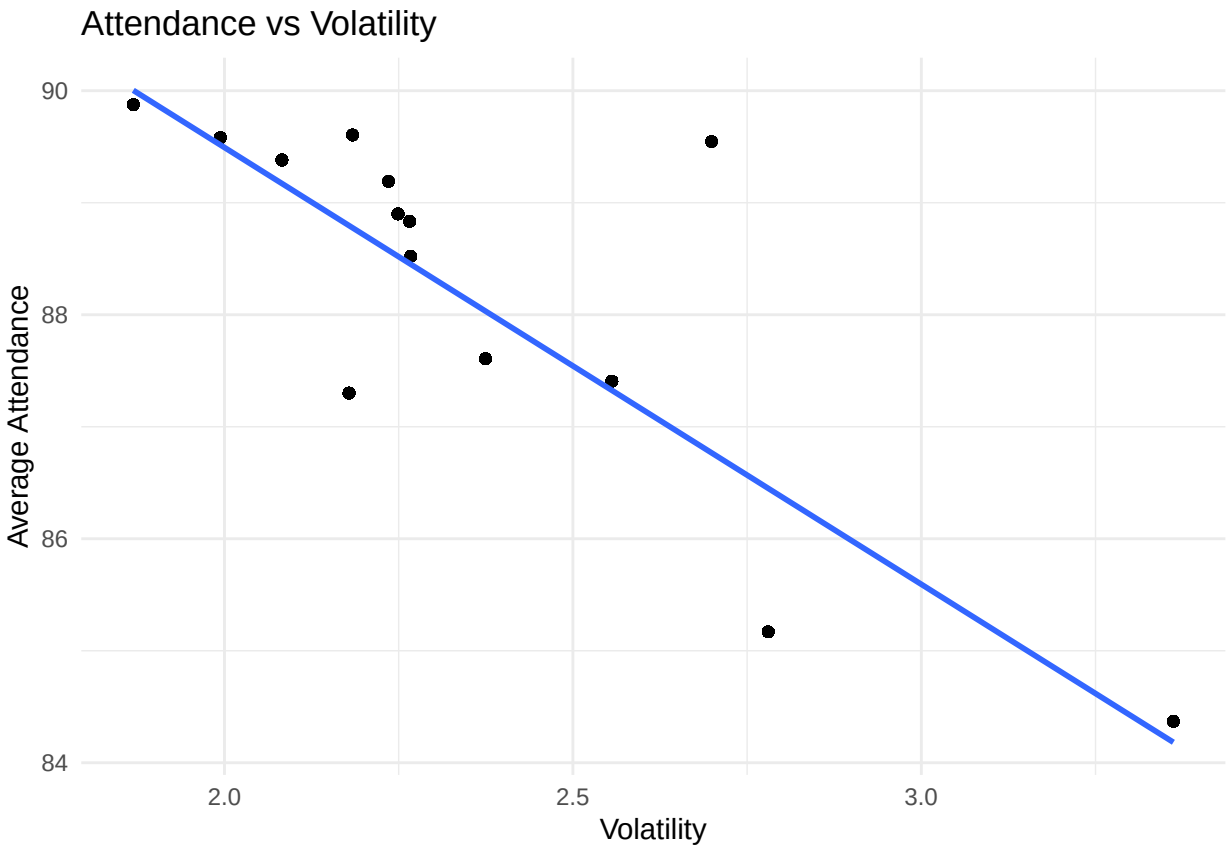
```
## # A tibble: 2 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept) 113.      0.353        319.        0
## 2 eqi_mean     -0.0526  0.000755     -69.6       0
```

```
## # A tibble: 3 x 5
##   term               estimate std.error statistic p.value
##   <chr>                 <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)        108.      0.144        750.        0
## 2 eqi_mean            -0.0268  0.000363     -74.0       0
## 3 volatility_present  -3.13    0.0254      -123.        0
```

```
## # A tibble: 4 x 5
##   term                         estimate std.error statistic   p.value
##   <chr>                           <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)                   73.9      1.41         52.5 0
## 2 eqi_mean                       0.0443   0.00295      15.0 3.38e- 49
## 3 volatility_present            12.0      0.625        19.3 8.98e- 78
## 4 eqi_mean:volatility_present   -0.0316   0.00130     -24.3 1.87e-118
```

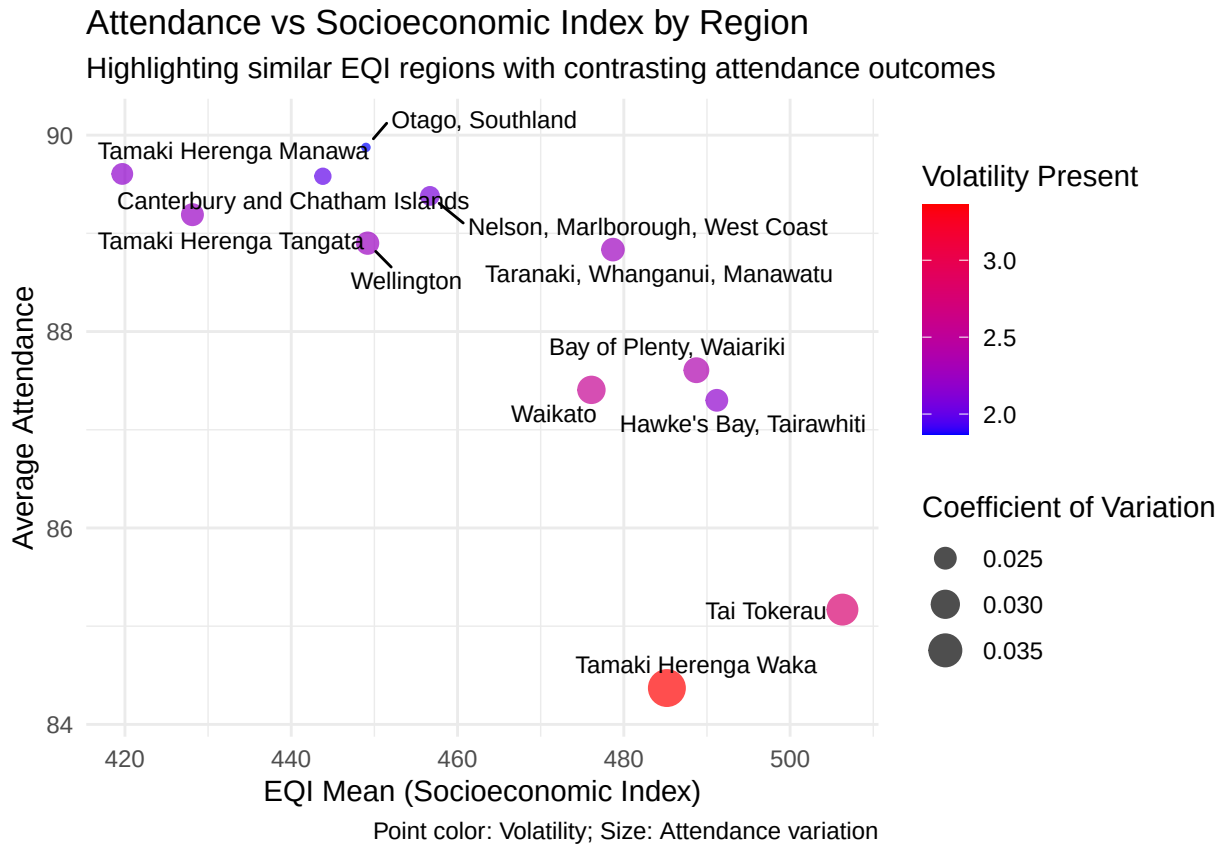
Visualising attendance vs volatility

```
## `geom_smooth()` using formula = 'y ~ x'
```



This scatter plot visualizes the relationship between attendance volatility and average attendance across regions. Each point represents a region's average attendance rate plotted against how volatile their attendance is over time. The linear trend line (fitted using linear regression) highlights the overall pattern in the data.

Volatility as a moderator of socioeconomic attendance risk



This analysis explores how volatility moderates the relationship between socioeconomic status (measured by the EQI mean) and average attendance across regions. The visualisation reveals that regions with similar socioeconomic profiles can have quite different attendance performances, influenced by volatility. It suggests that attendance volatility acts as a moderator, affecting how socioeconomic risk translates into actual attendance outcomes.