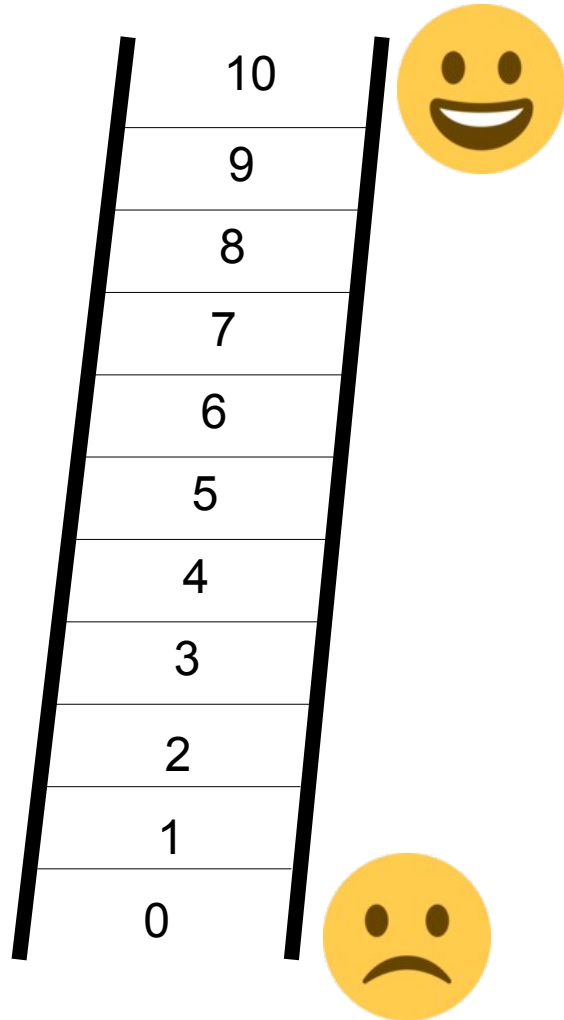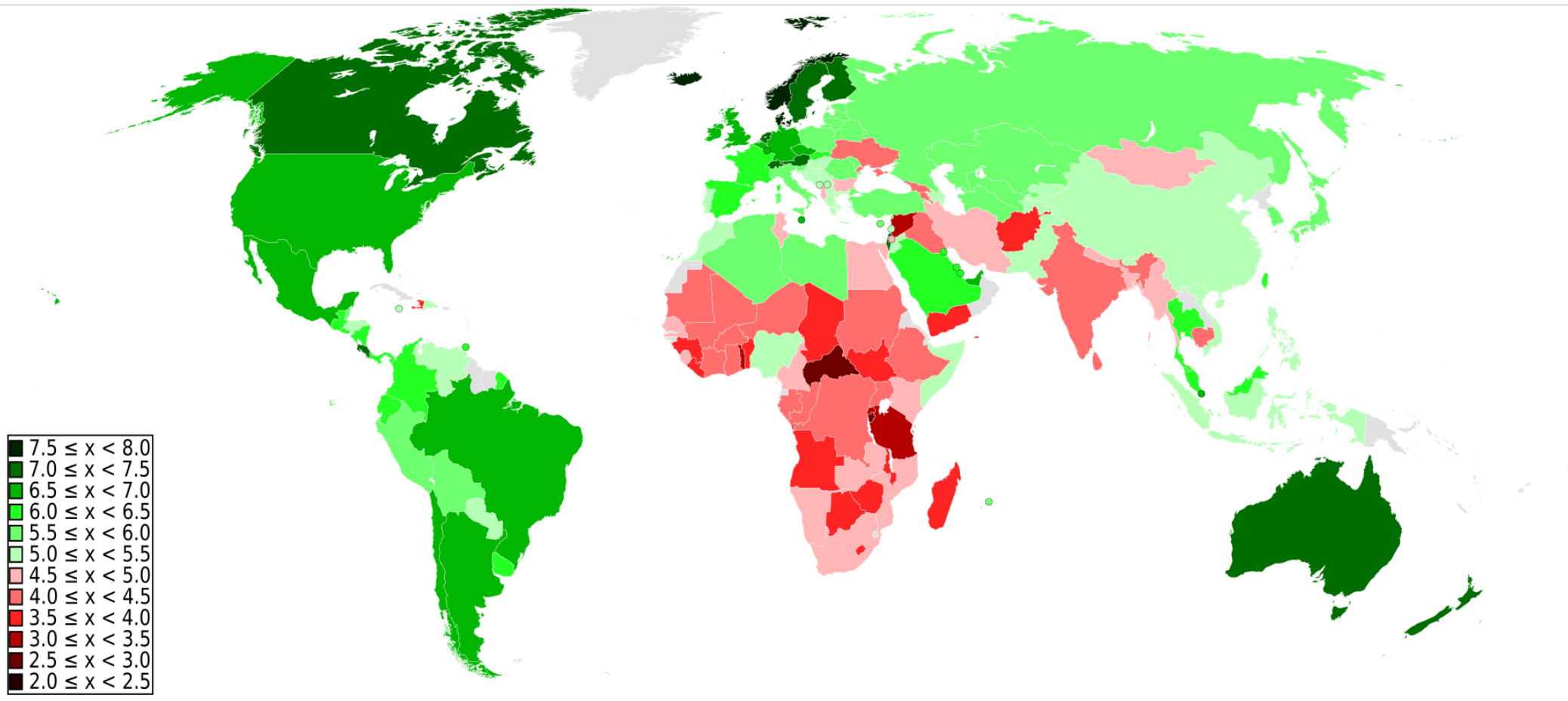# The Pursuit of Happiness

Thomas Anderson, Peter Hanrahan, Roman Maron, Alan Treanor, Peng Ye, Jiali Yu

The Problem

- Cantril Ladder
- Not a calculated value

- What other factors influence Happiness?
- Can we predict it?

Legend:
- 7.5 ≤ x < 8.0
- 7.0 ≤ x < 7.5
- 6.5 ≤ x < 7.0
- 6.0 ≤ x < 6.5
- 5.5 ≤ x < 6.0
- 5.0 ≤ x < 5.5
- 4.5 ≤ x < 5.0
- 4.0 ≤ x < 4.5
- 3.5 ≤ x < 4.0
- 3.0 ≤ x < 3.5
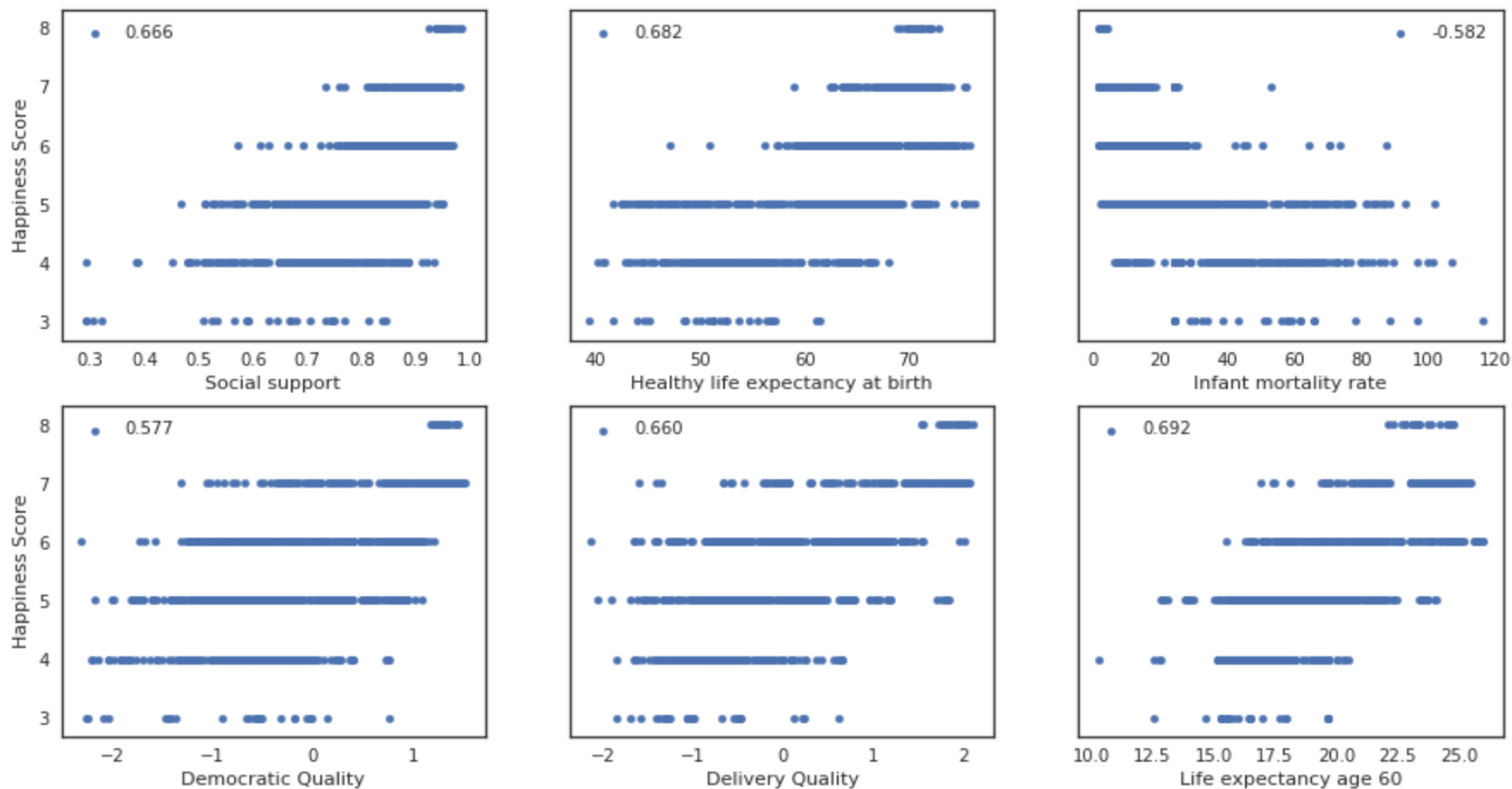- 2.5 ≤ x < 3.0
- 2.0 ≤ x < 2.5

The Data

- 10 years
- *Many* features – difficult to choose
  - 20 descriptive features

- Combining datasets proved difficult
- Different sources:
  - Year ranges

  - Frequencies

  - Overlap between datasets
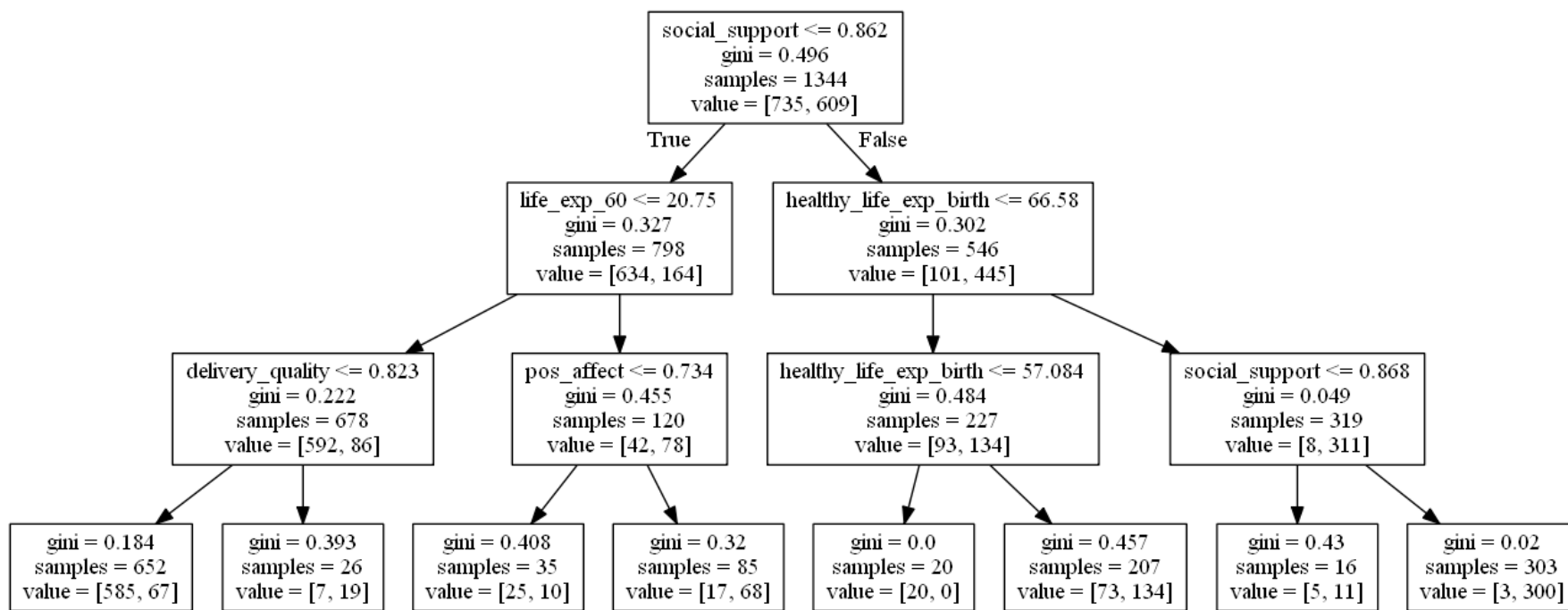
The Analysis

Model – Linear Regression

- Initial Results
  - R^2: 0.716
  - P-values: All chosen features within 95% confidence
- Normalisation
  - No improvements
  - Highest co-efficients: *Social Support & Life Expectancy at Age 60*
- Standardisation
  - R^2 dropped to 0.683
  - Highest co-efficients: *Social Support & Healthy Life Expectancy at Birth*

Model – Logistic Regression

- Accuracy:
  - 82% on Training set
  - 83% on Test set
- Cross-validation
  - Accuracy: 82%
  - F1-score: 80%
  - Possible "lucky shuffle"

Model - Tree/Forest

- Optimal Results at depth 3

- Accuracy: 85%
- Precision: 86%
- F1-score: 86%

Conclusion

- Long life
- Access to support in crisis
  - Healthcare
  - Welfare

Bonus Slide

Trend of Happiness Score Worldwide