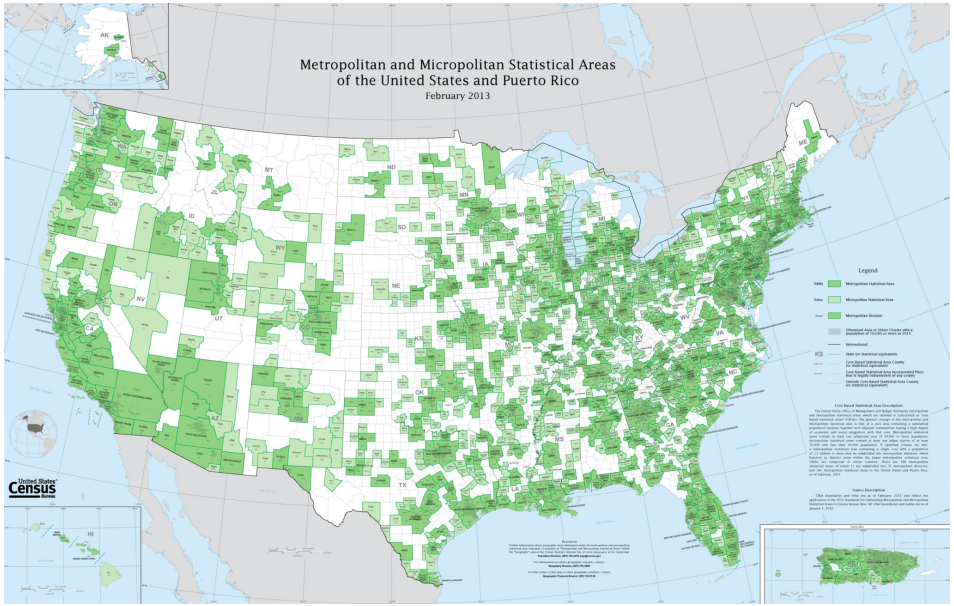
Walkability vs. Subjective Well-Being

Group 15: Robert Jiang, Josna John, Collin Jun, Justin Kane-starr, and Ziyan Zhou

Walkability Dataset Introduction

Data Collection Method

- Environmental Protection Agency
  - GTFS data
- Core Based Statistical Areas
  - CBSA
- Over 100 features and 220,000 points



The ranked scores were then weighted by the following formula:[16]

$$\text{Final National Walkability Index score} = (w/3) + (x/3) + (y/6) + (z/6)$$

Where w = block group's ranked score for intersection density

x = block group's ranked score for proximity to transit stops

y = block group's ranked score for employment mix

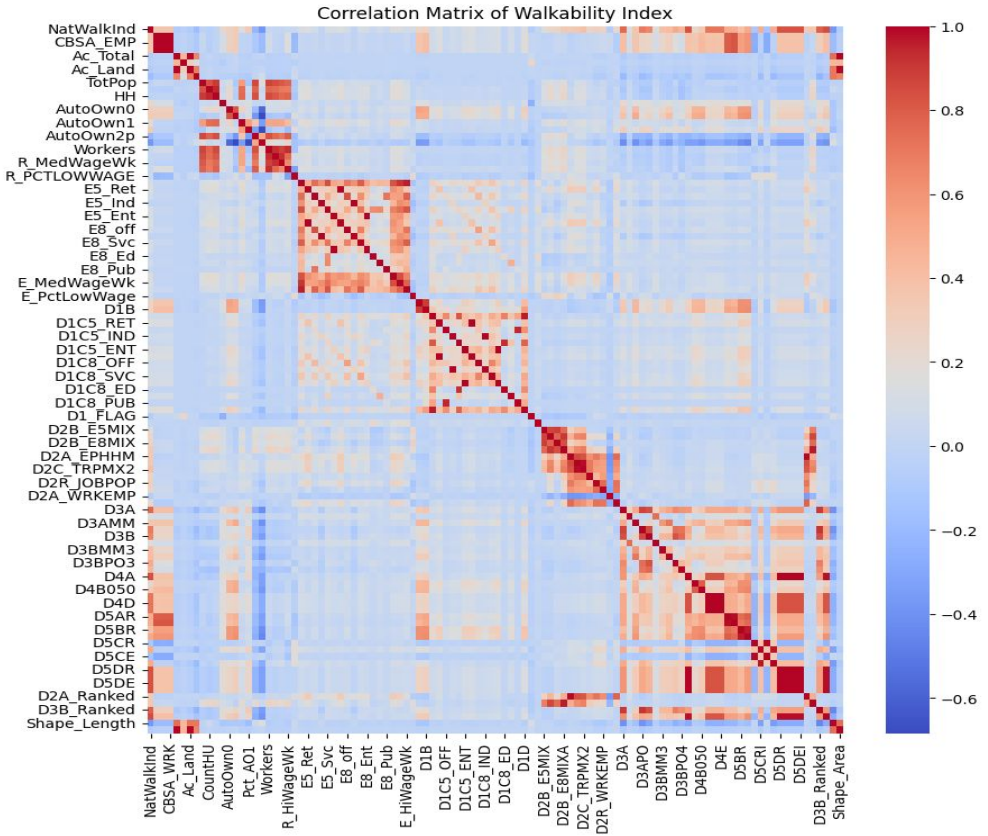
z = block group's ranked score for employment and household mix

The block groups are assigned their final National Walkability Index scores on a scale of 1 to 20 (Figure 3). The scores are categorized as follows:

	1 – 5.75	Least walkable
	5.76 – 10.5	Below average walkable
	10.51 – 15.25	Above average walkable
	15.26 – 20	Most walkable

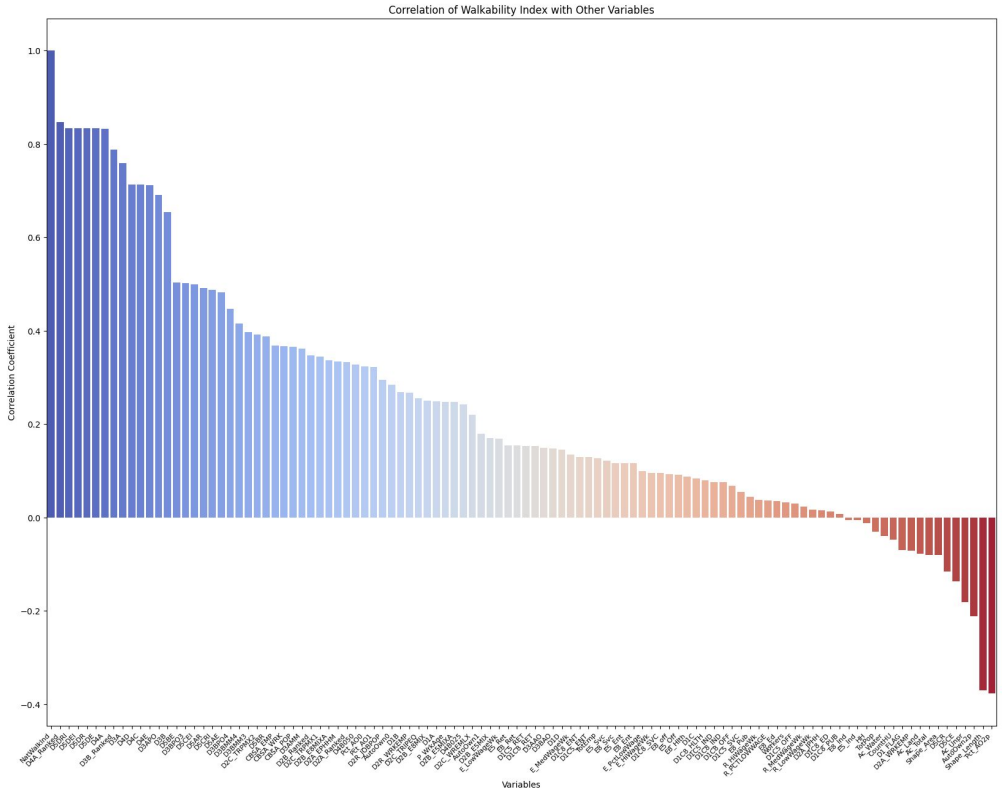
Motivation for dimensionality reduction

- Related features examples
  - Population of area, housing units
  - Households that own 1 car, 2 cars
- Correlation matrix indicating correlations between variables
- Red squares indicate high correlation



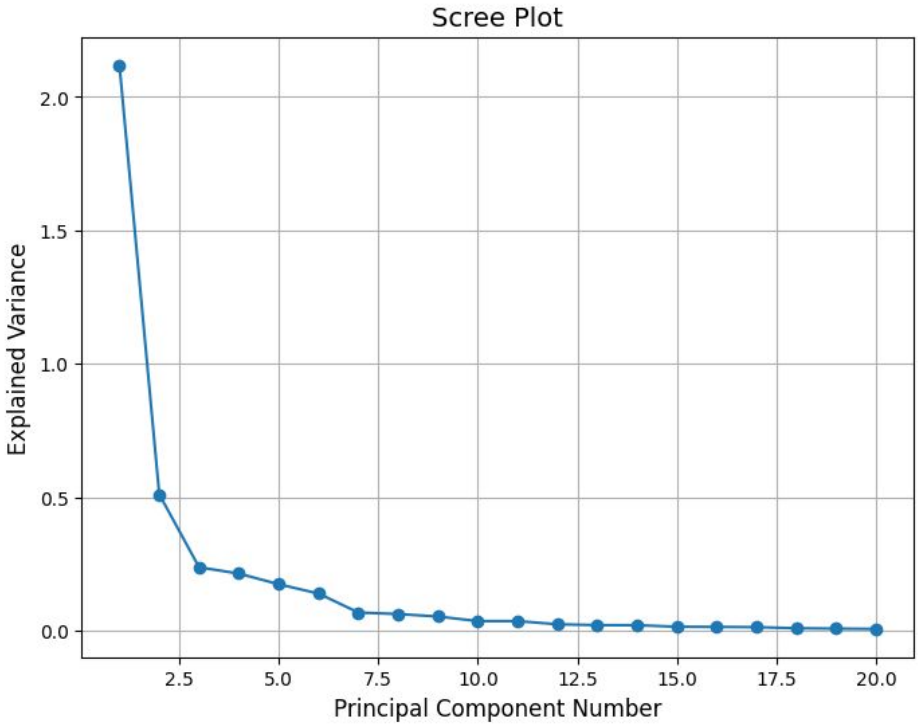
Walkability vs. Geospatial Metrics

- Graph displays correlation of variables with the current walkability index
  - Few variables have higher than 0.8 correlation coefficient
- Weights could be adjusted to simplify and/or improve walkability index

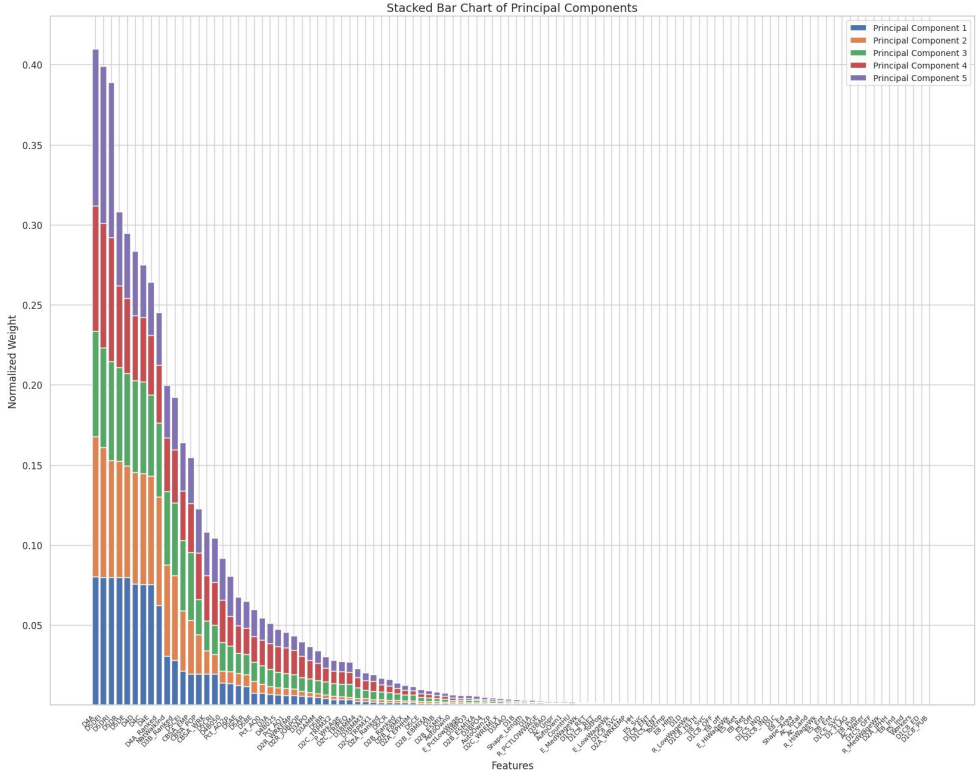


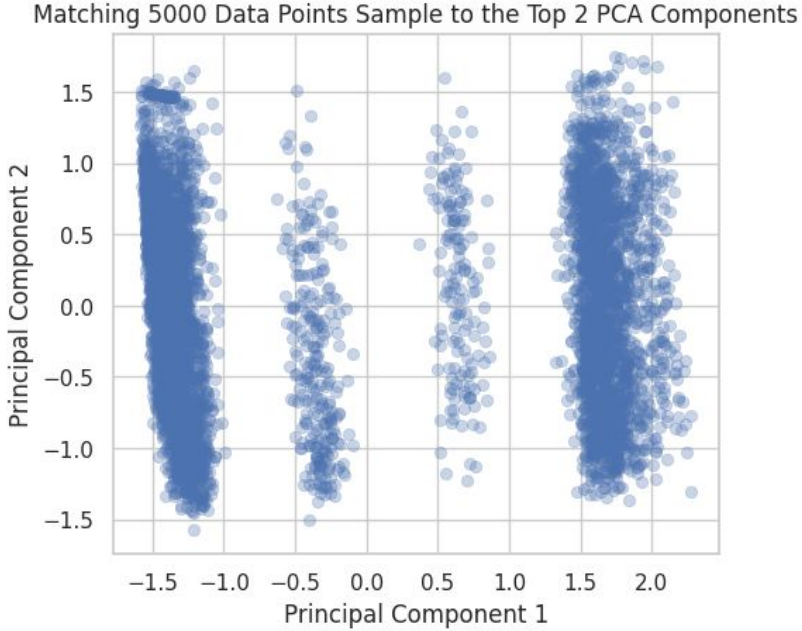
Dimensionality Reduction

- Principal Component Analysis
  - Most of variance is captured in a few components
  - Can represent most information with a very low dimension from over 100



Dimensionality Reduction





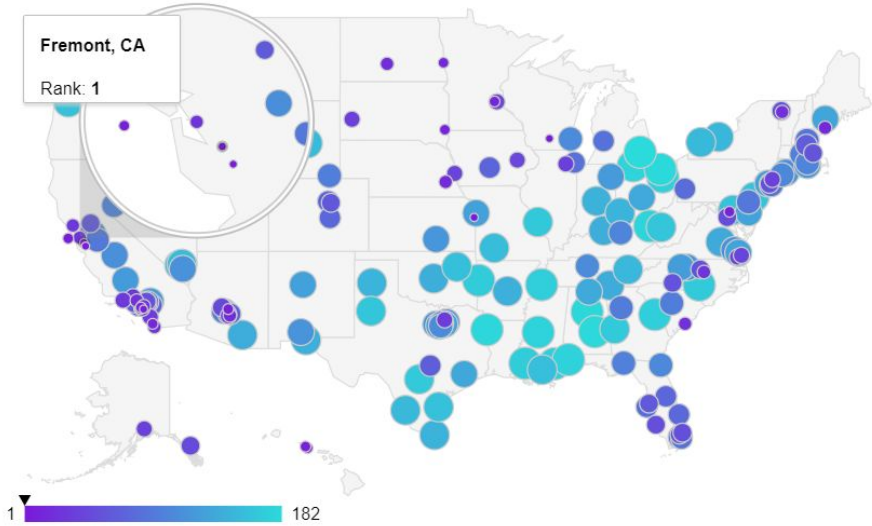
Methodology for New Walkability Index

- Several ways to create a new walkability index algorithm:
  - A PCA-inspired index where the top PCA components are used to define the new index algorithm
  - Out-source the data to define a happiness-based walkability index algorithm

Happiness Index Dataset Introduction

Data collection method:

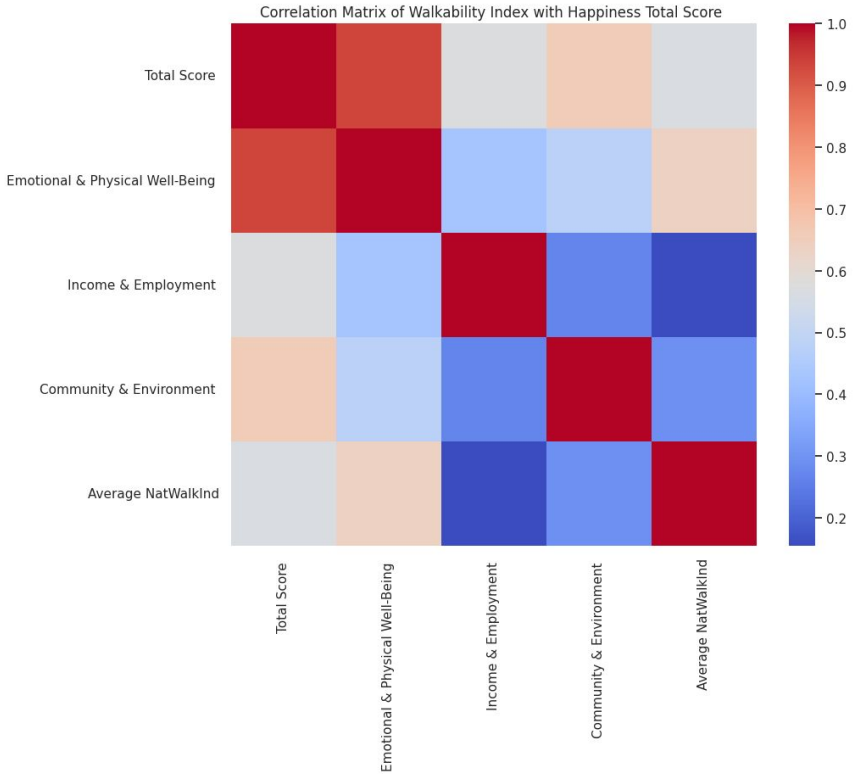
- Information collected at the 182 largest cities in the United States
- Top weighted components include
  - Life-satisfaction index
  - Depression rate
  - Suicide rate



Overall Rank	City	Total Score	Emotional & Physical Well-Being Rank	Income & Employment Rank	Community & Environment Rank
1	Fremont, CA	74.16	1	85	3
2	Overland Park, KS	69.78	6	36	6
3	San Jose, CA	69.54	2	33	37
4	Madison, WI	68.99	7	17	9
5	Irvine, CA	68.22	4	26	24

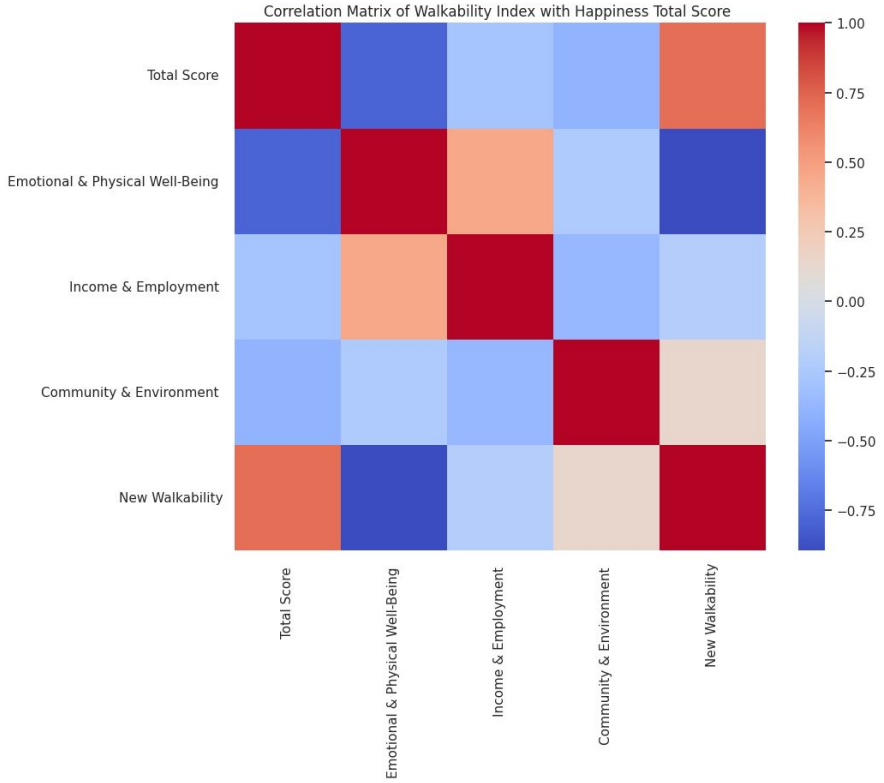
Walkability vs. Happiness Correlation

- Correlation between Happiness and walkability is quite reasonable (0.56)
- Index algorithm can be improved to show correlation between walkability and happiness



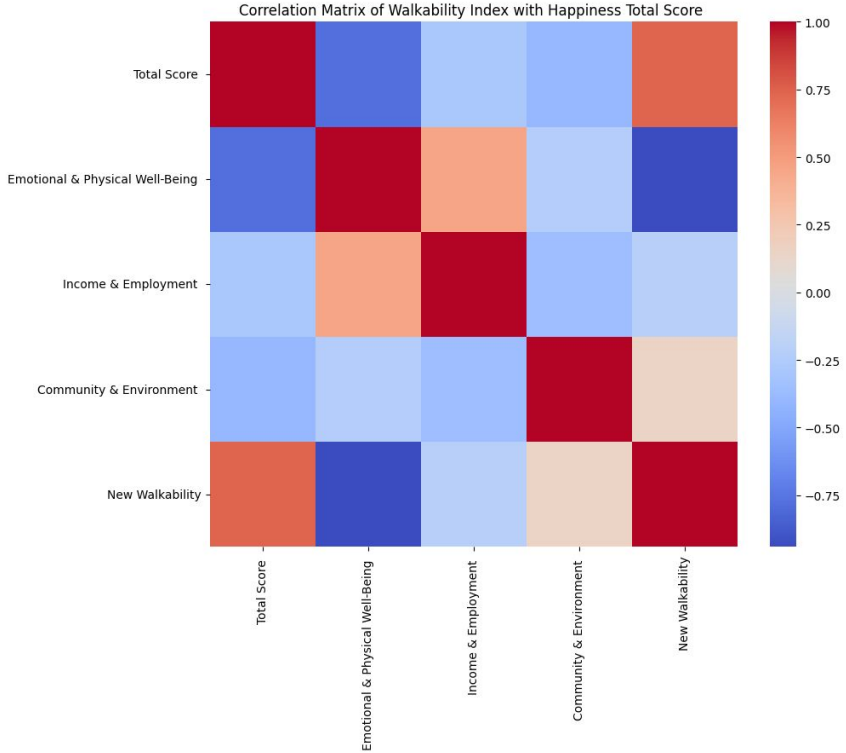
New Walkability Index - Linear Regression

- Using a linear regression model, we were able to create a new walkability index
- The correlation coefficients greatly differ when compared to the first
    - Higher correlation between the new walkability and total happiness score



New Walkability Index - Gradient Boost Regression

- Using a gradient boost regression model, we were able to achieve very similar results to that of the linear regression model
    - Have tried other models such as random tree regression and random forest regression and the results were very similar
- The walkability index has been improved and the correlation coefficient to the happiness index has been optimized.



Conclusions

- Using dimensionality reduction, the top components were identified and used to help define the new index algorithm.
- The original walkability index and the happiness index was shown to have a positive correlation, showing that walkability has a positive impact on the happiness within a community.
    - The happiness index and its variables were used to define the new walkability index.
- Walkability index was changed/improved using a PCA-inspired index and now displays a stronger correlation with the happiness index.
    - Can possibly be further improved, but major improvements were made
- Improving this walkability dataset can enhance the effectiveness of urban planning decisions.

Future Scope

- Recommender System
  - An algorithm to determine a person's ideal city to live in based on an individual's weighted preferences (e.g., walkability, income)
- More visualization based on geographic data

References

- https://catalog.data.gov/dataset/walkability-index1
- https://wallethub.com/edu/happiest-places-to-live/32619
- https://en.wikipedia.org/wiki/Core-based_statistical_area
- https://www.epa.gov/smartgrowth/smart-location-mapping#Trans45

Code Repository

https://github.com/zhou-ziyan/ECE143Project