

# American Life

## An Embedding Study of the ATUS Dataset

---

Yuan Chen

chen2243@wisc.edu

Department of Computer Science  
University of Wisconsin - Madison



# Introduction

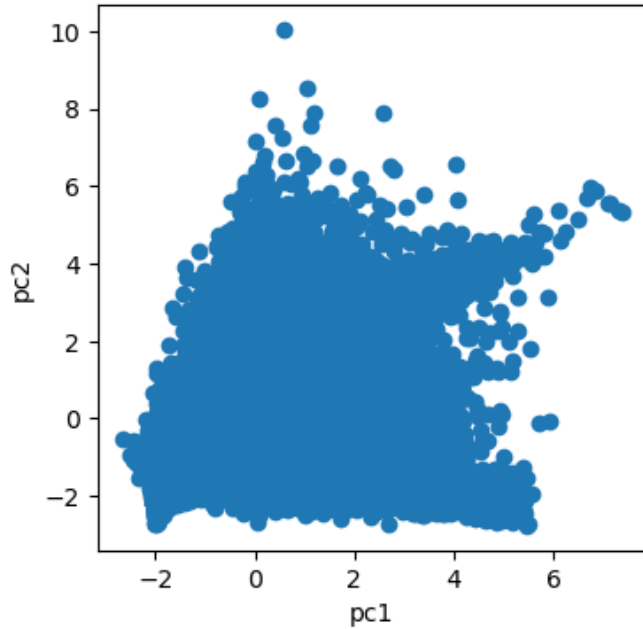
---

- ❖ Purpose: Use embedding to find out what factors determines the time usage of an average American
- ❖ Tools: [Python](#)(Pandas, UMAP, Matplotlib), [Tableau](#)
- ❖ Data: ATUS

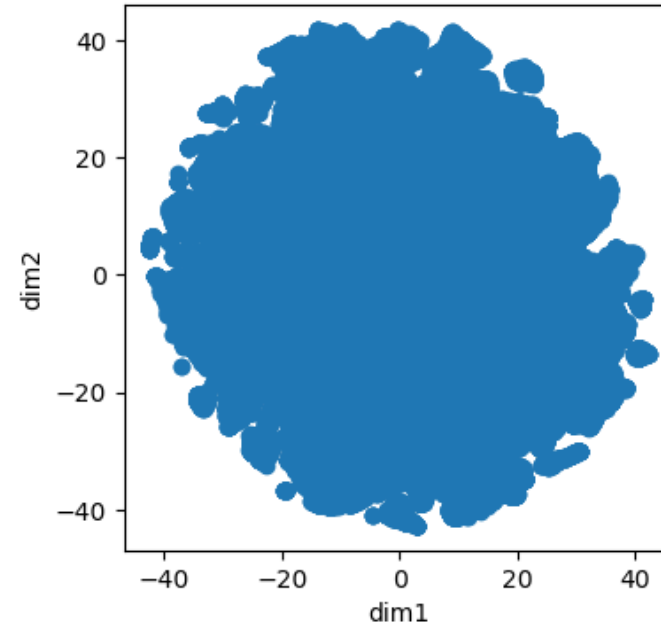


# **Compare Dimension Reduction Methods**

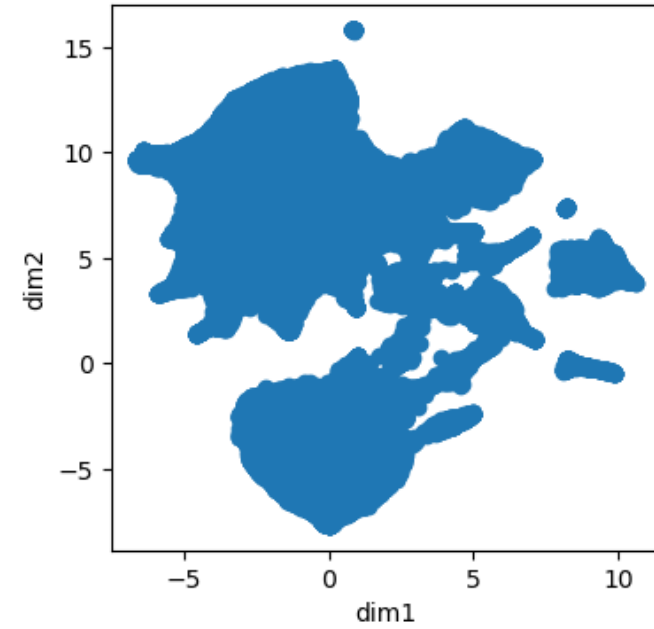
# Embedding Techniques



PCA



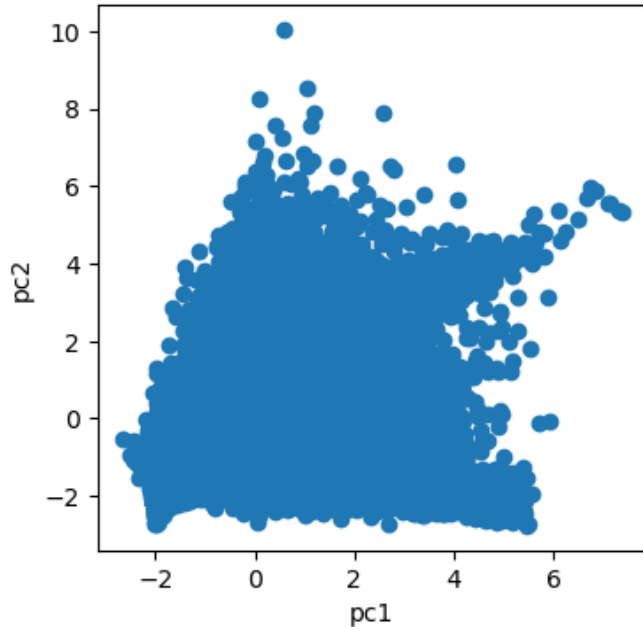
t-SNE



UMAP

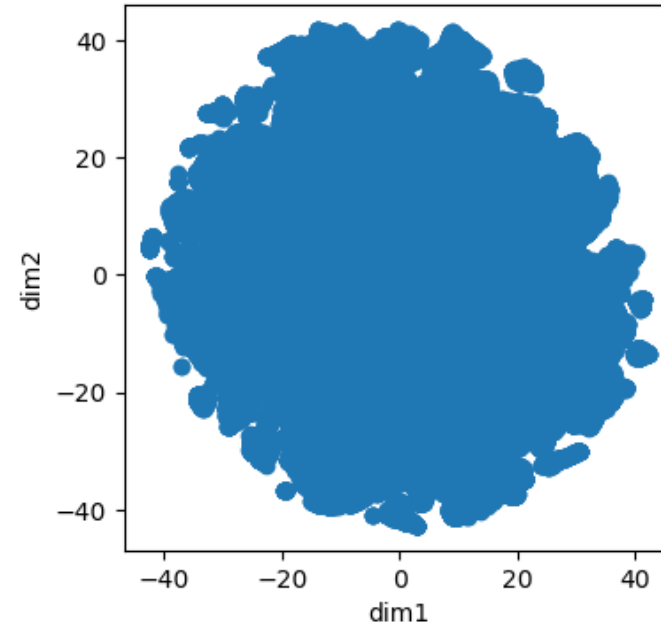
Difference?

# Embedding Techniques



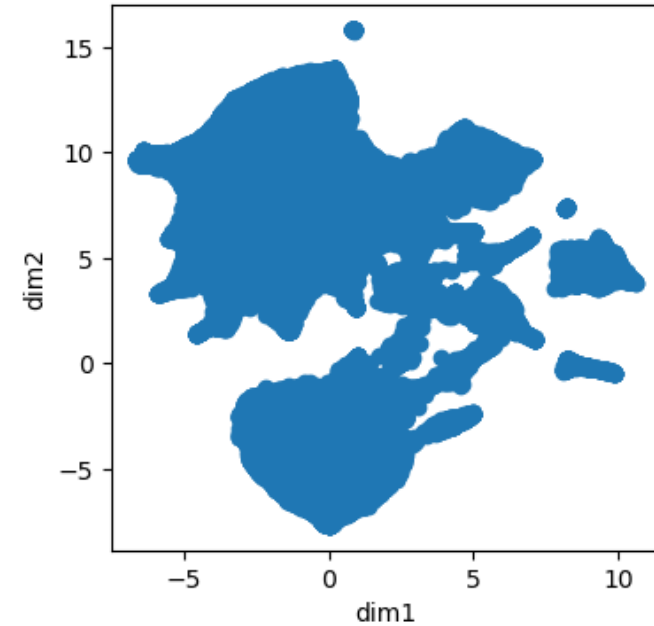
PCA

Based on most significant  
variants, information lost during  
the reduction



t-SNE

Slow  
Need tuning



UMAP

Fast  
Preserve more info  
Good clustering

# Find Patterns



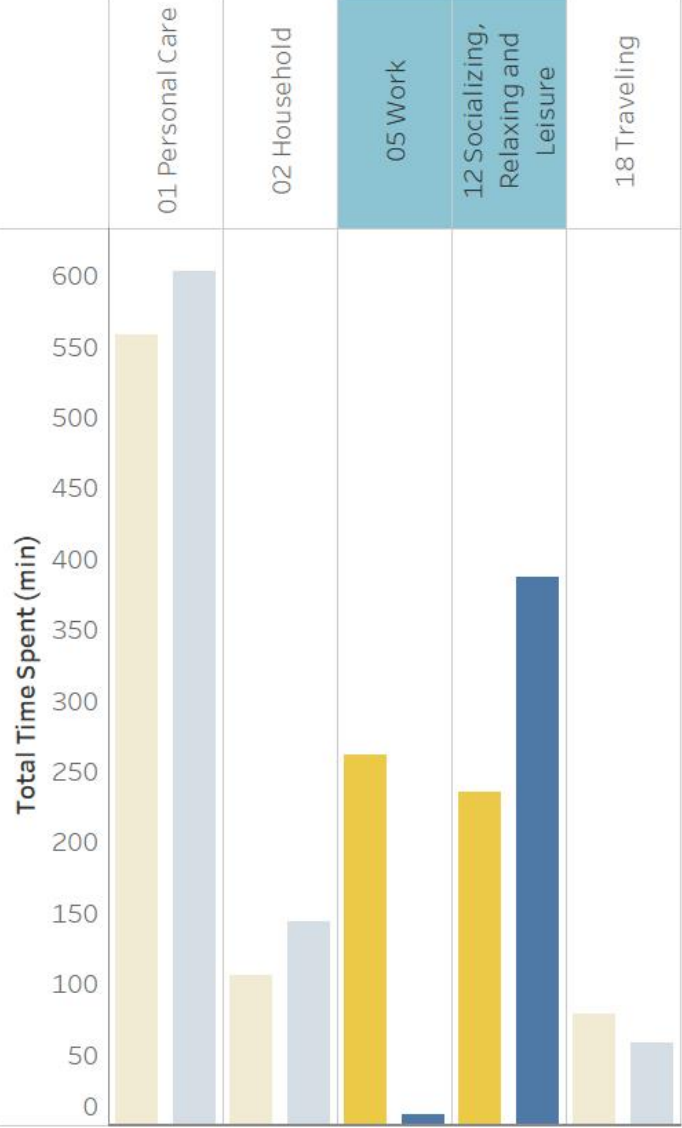
# Clusters

clustering based on employment

■ Unemployed or Absent  
■ Employed



time usage based on employment

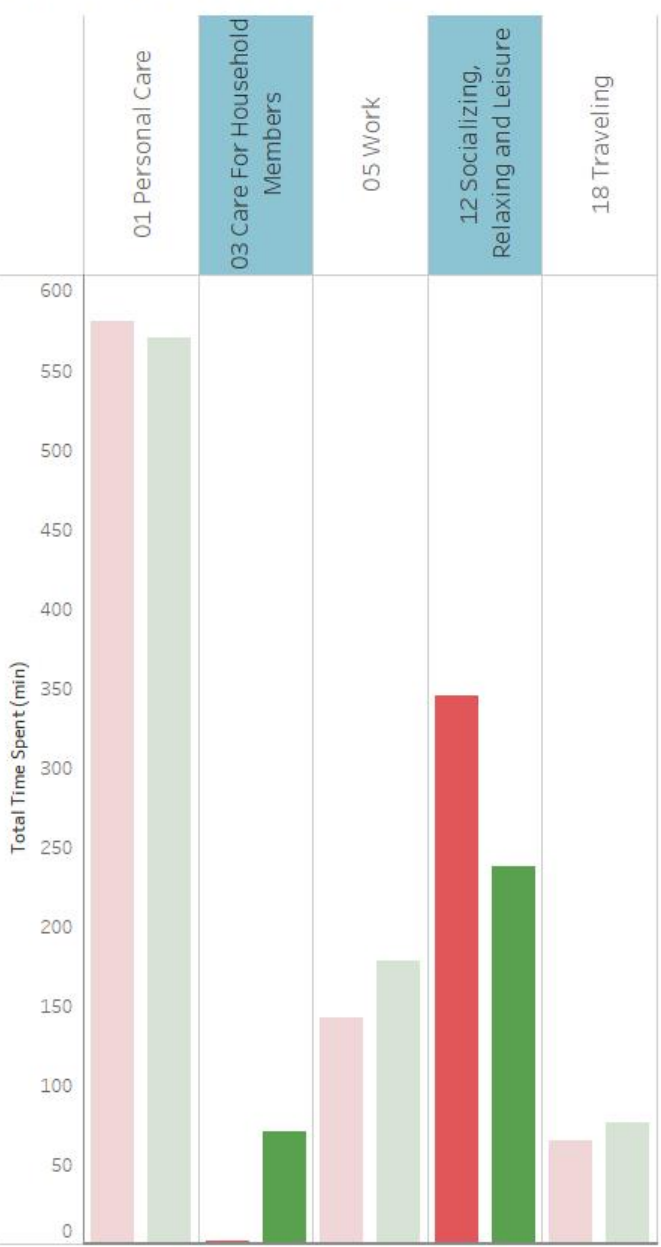




clustering based on having children



time use based on children



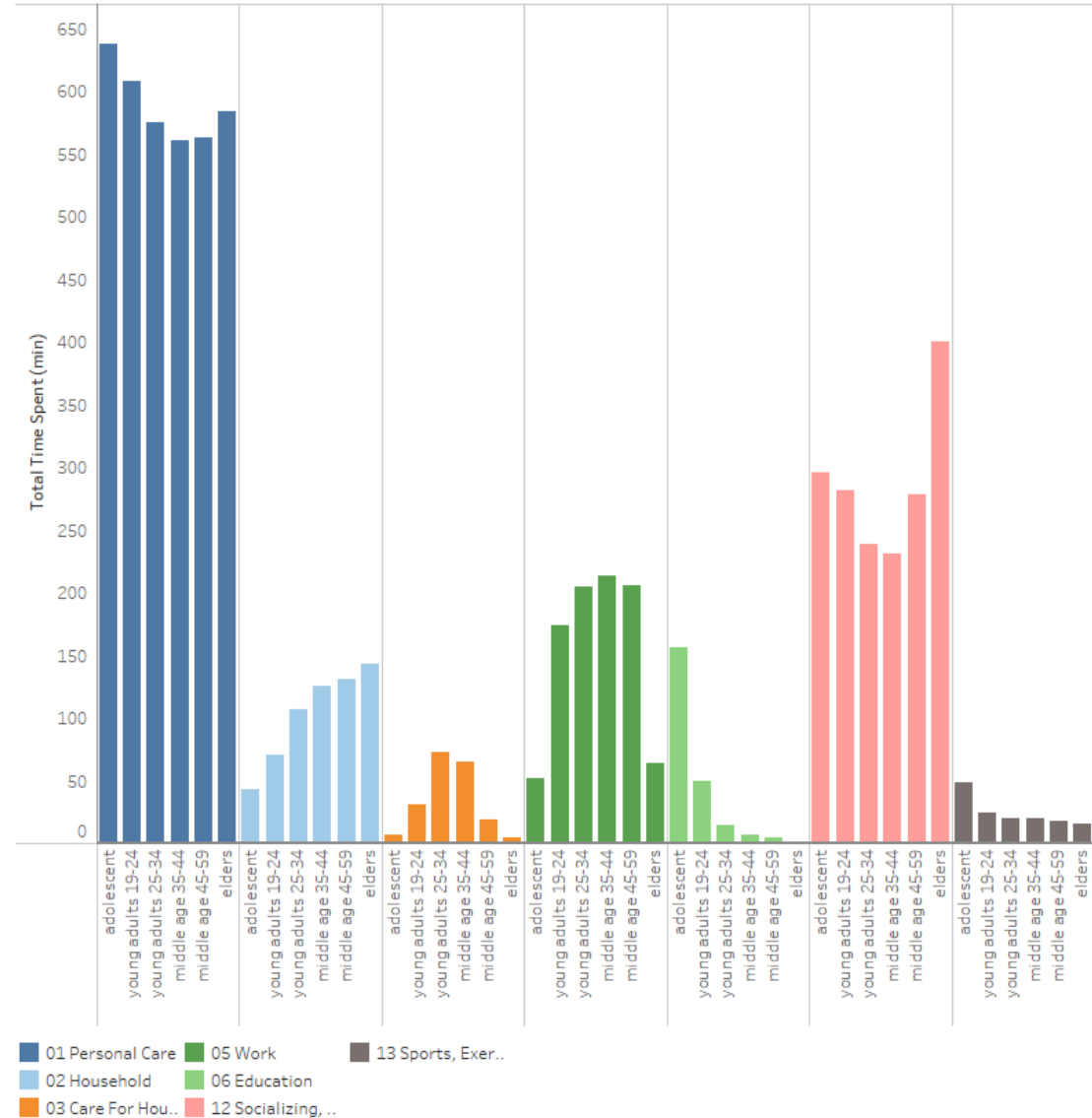


# Life of An Average American

clustering based on age groups



time usage based on age group



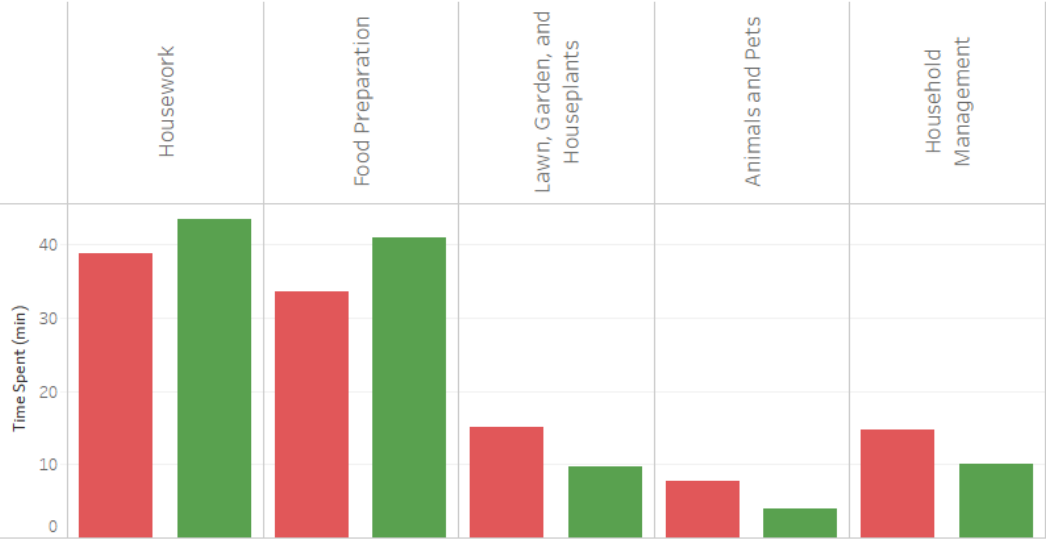
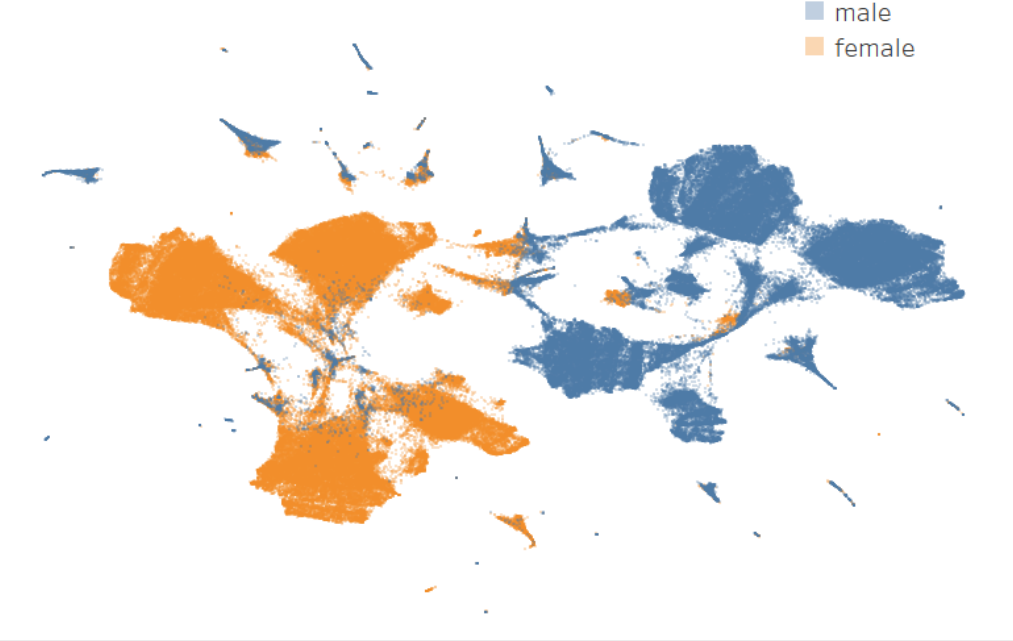


# **Validate Previous Data Exploration**

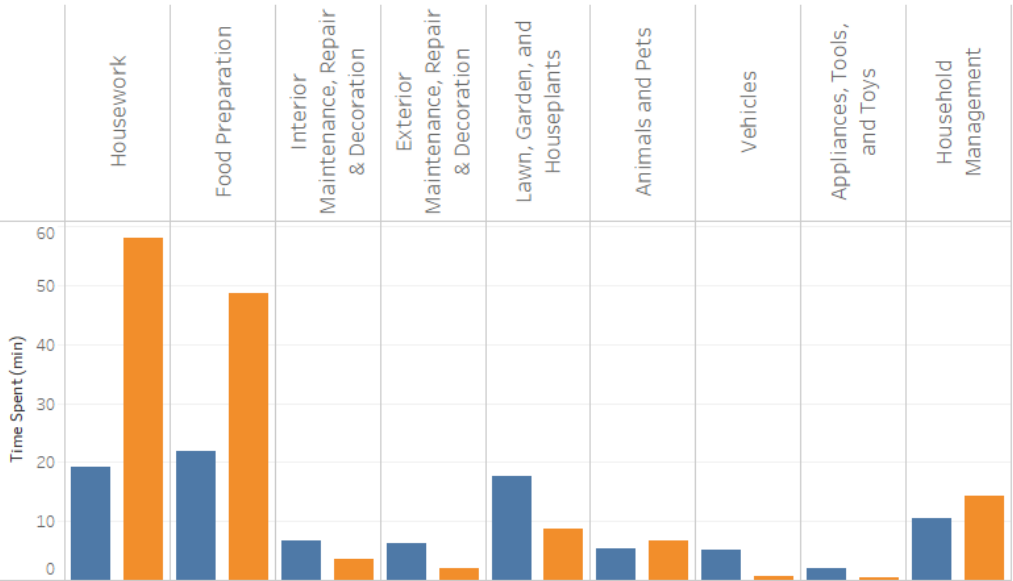
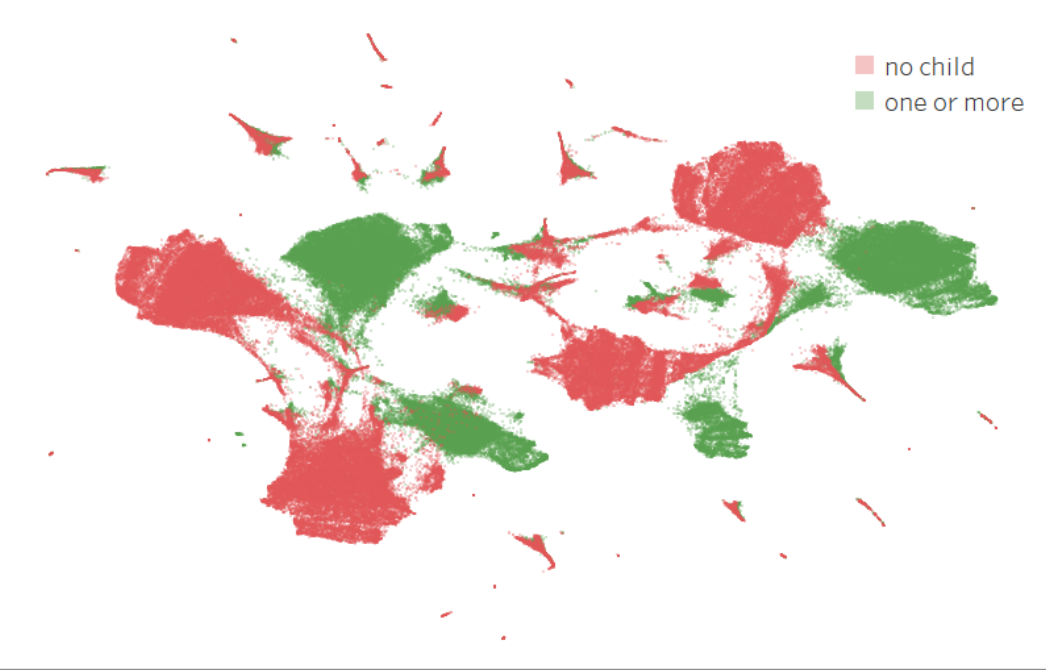
# Gender Gap in Household Activities



time spent on household activities by gender



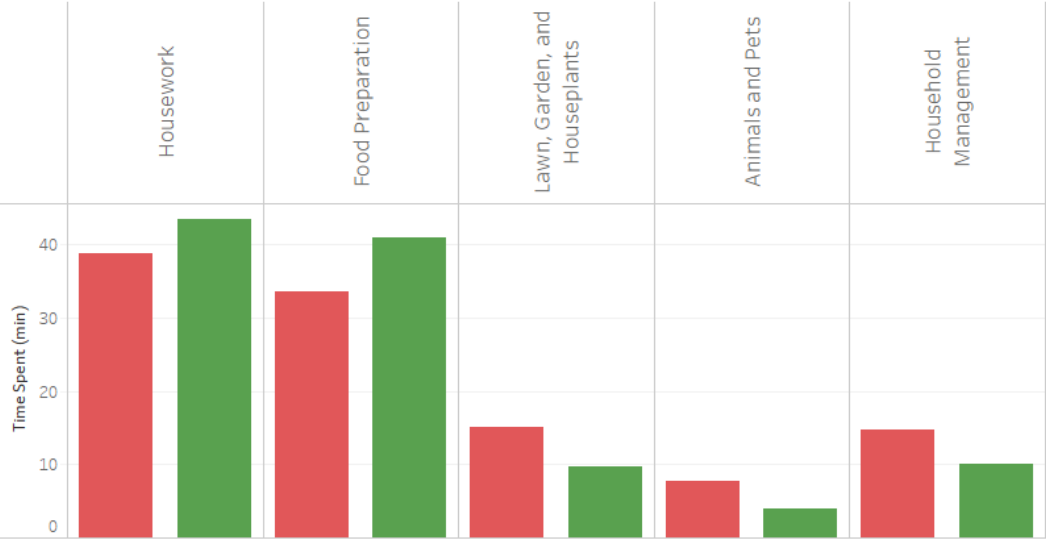
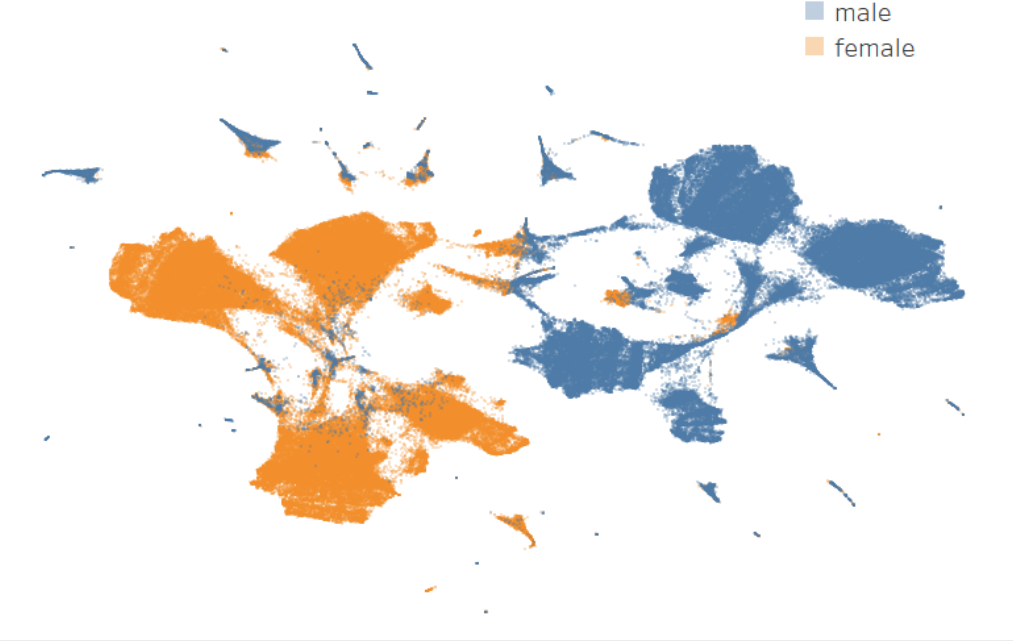
time spent on household activities by having children or not



# Gender Gap in Household Activities



time spent on household activities by gender



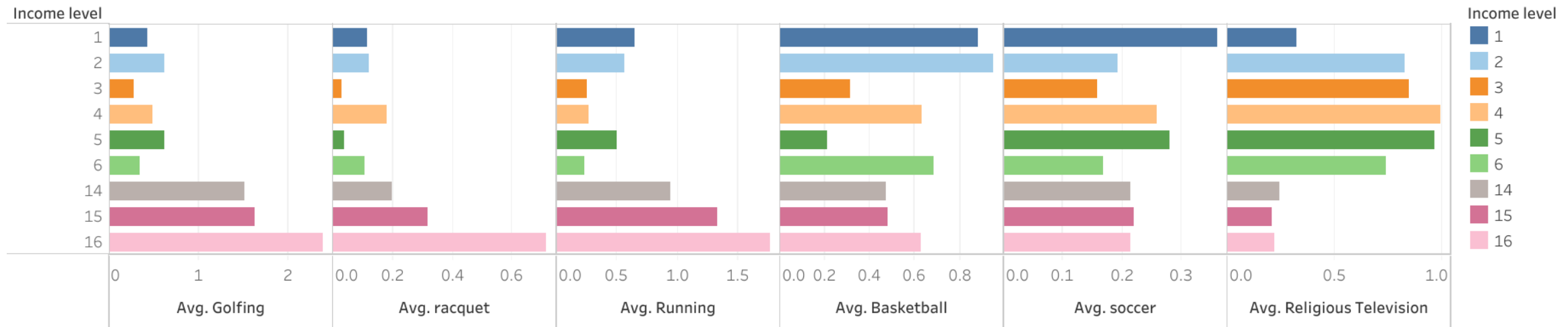
time spent on household activities by having children or not



# Previous Critique – What Sports Do Wealthy People Play

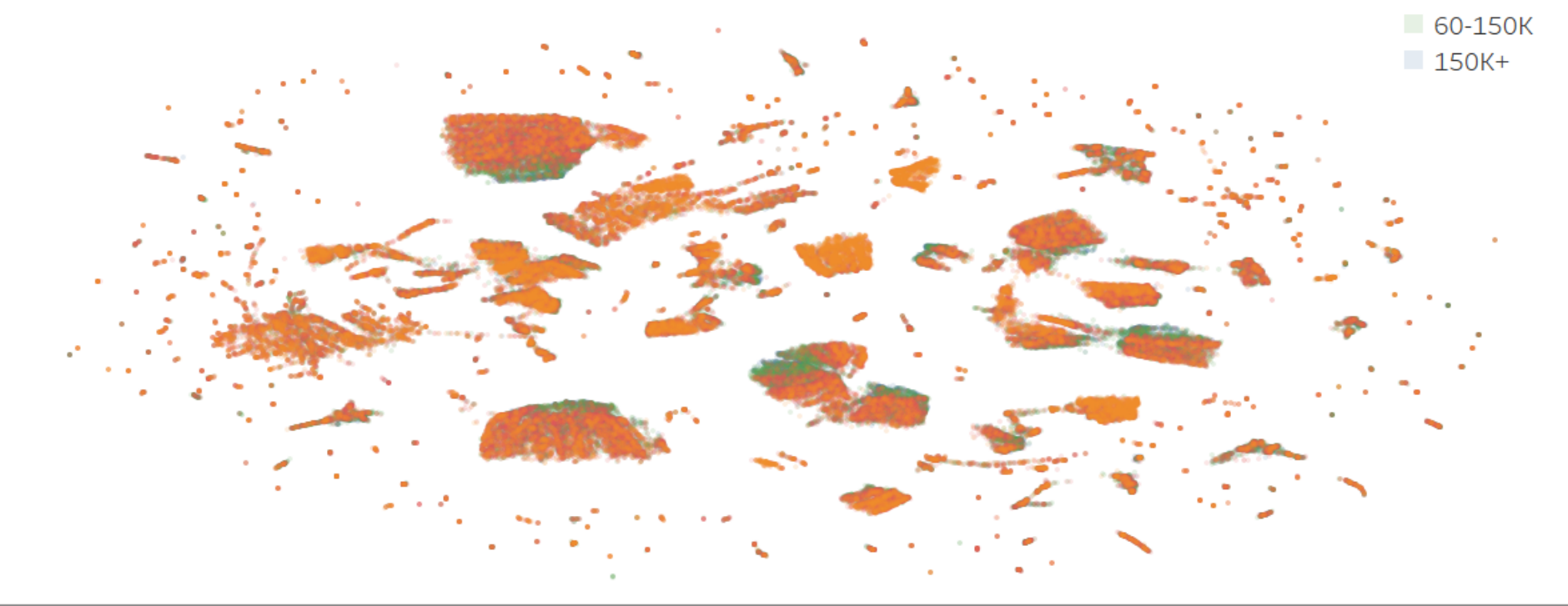


Wealthy Sports

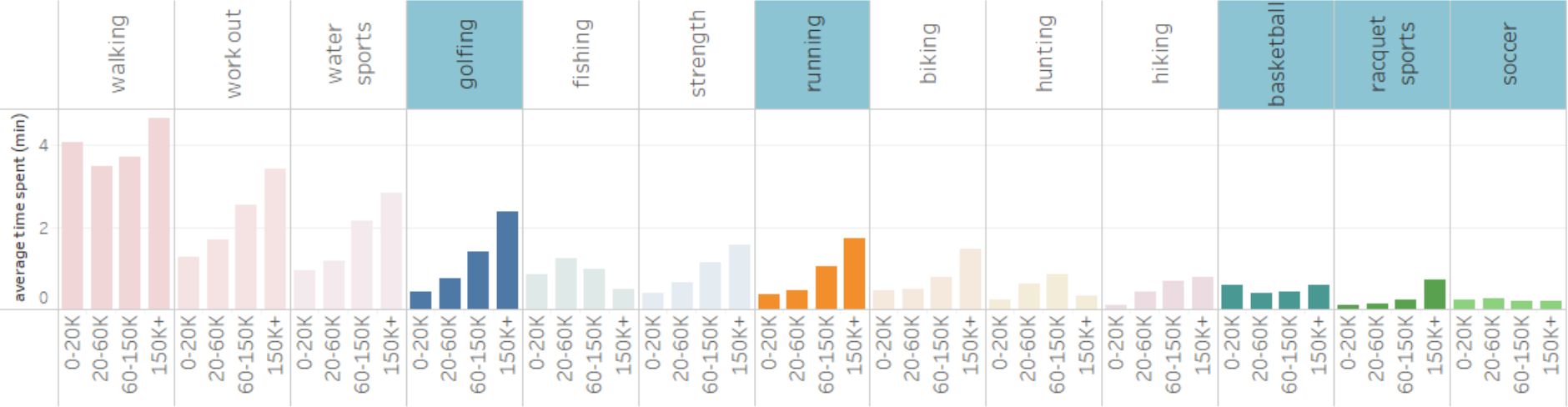


From #37 in the critique exercise  
Is it true though?

time spent clustering based on income



total sport time ranking

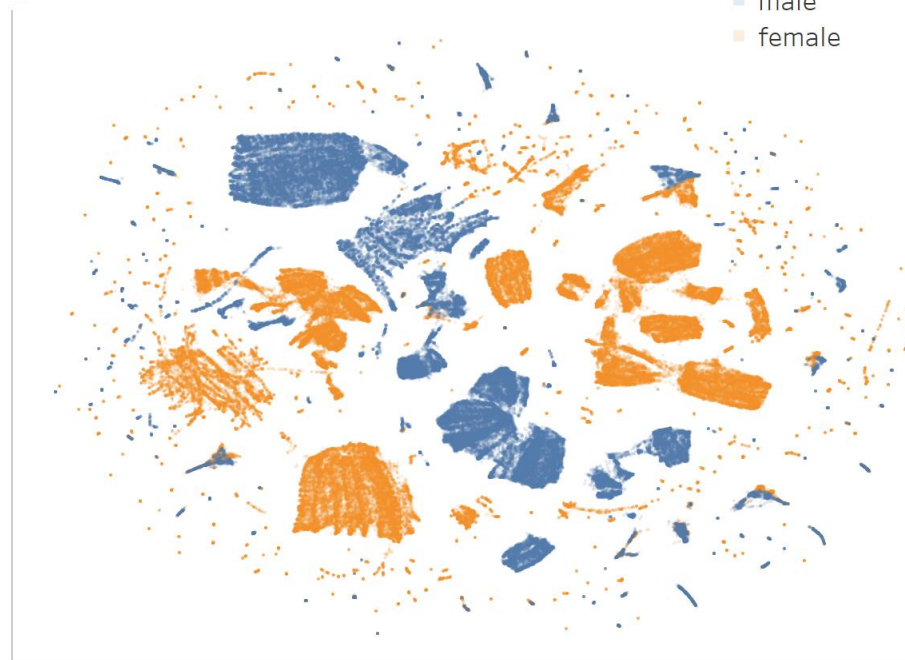




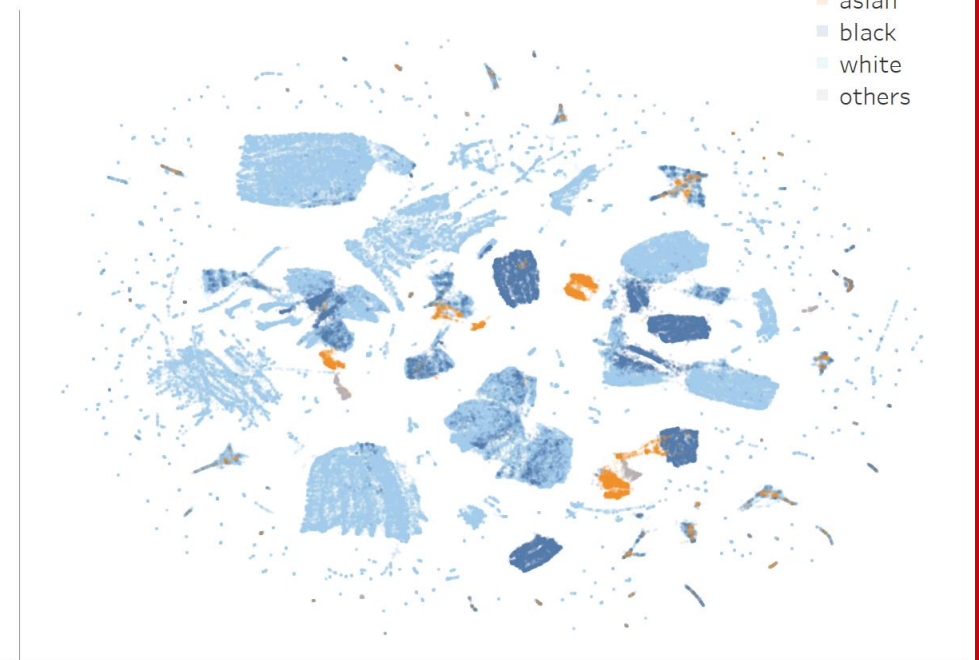
# Real Factors?

- Gender
- Race
- Children

gender



race



has children or not



education





# Summary

- ❖ Compared different dimension reduction methods
- ❖ UMAP embedding showed the life of an average American
- ❖ Embedding can be used to validate patterns
  - Confirmed the gender gap in household activities
  - Found more impactful factors that determines who play what sports

