

# 作业报告

1800011307 李星桥

# 数据描述与分析

数据集 `data.csv` 为来自27所世界知名大学的教职人员的相关数据，表现为表格形式，维度为14，主要包括Institution、Name、PhD Graduate Year/School、Research Interest、H-index、Publications等数据信息。

数据集各维度不是独立的。例如，Institution和Institution Index为一一对应关系。除此以外，注意到数据维度PhD Graduate Year和H-index、Publications、Citations、Publication Divided by Co-authors等数据为连续数值型数据，考虑利用这些数据进行可视化展现。

另外，数据集中有一些隐含的关联，通过可视化，我们可以找出这些关联。例如，毕业年份与科研水平的关联；毕业学院与就职机构的关联；某机构的整体科研水平评估；某机构研究人员的毕业年份分布；等等。在本次可视化项目中，我们集中于毕业年份与科研水平的关联，以及机构整体科研水平的可视化。