

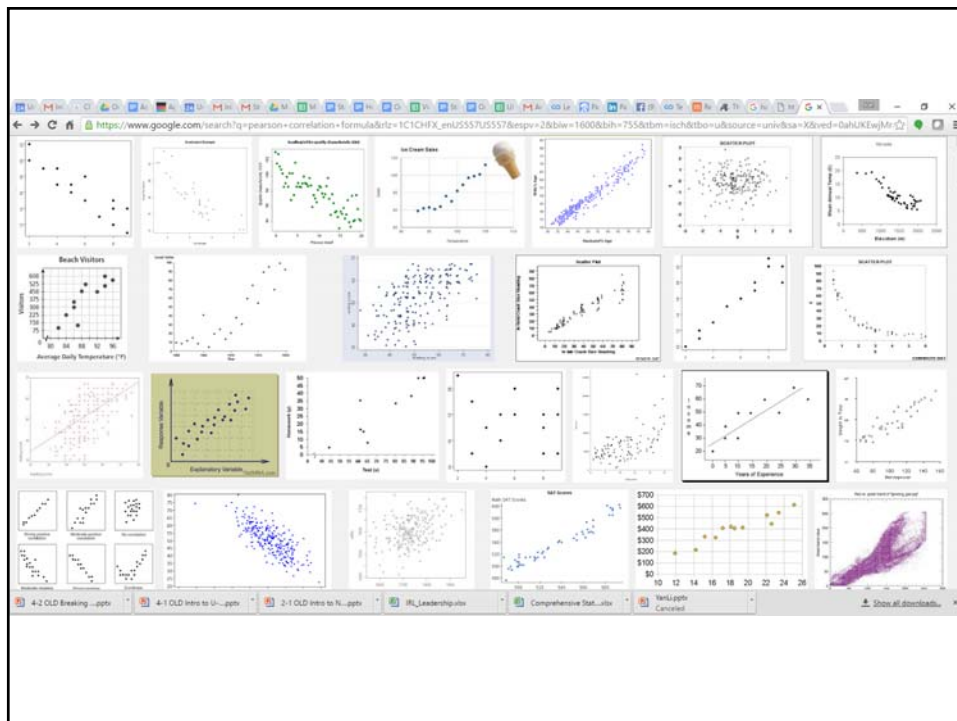
Demographics and Related Approaches

Demographics: What and Why?

- Motivation
 - Popularity may not represent my tastes
 - I may be part of an identifiable cohort with different tastes
 - Age
 - Gender
 - Race/Ethnicity
 - Socio-Economic Status
 - Location
 - Including non-demographics that may be predictive

OK, But How?

- Start by identifying available demographics (and correlates)
 - Many will require processing or bucketing
 - Age is often divided into groups
 - Postal codes can be transformed into socio-economic status, urban/rural, dominant ethnicity, etc.
- Then explore where your data correlates with demographics
 - Scatterplots, correlations,



If You Find Relevant Demographics

- Step 1: Break down summary statistics by demographic
 - E.g., most popular item for women, for men
 - Maybe even factorial (most popular item for men 45-60)
- Step 2: Consider a multiple regression model
 - Predict items based on demographic statistics
 - Linear regression for multi-valued (e.g., rating) data
 - Logistic regression for 0/1 (e.g. purchase) data

Important!

- You need defaults for unknown demographics
 - May simply be overall preferences
 - May reflect expected demographics of newcomers
 - May be modeled separately
- If demographics are useful, getting data on users is key
 - Various sources of data, from advertising networks to loyalty club sign-ups and surveys
 - In some cases, demographics can be “predicted” from data.

The Power and Limits of Demographics

- In many cases, demographics work because products or content is created to reach them:
 - Television programs
 - Magazine articles and advertisements
 - Personal products
- Or products simply naturally appeal to different groups
- Of course, demographics fail miserably for people whose tastes don't match their demographics!

Demographics and Related Approaches