

DS 1300 Lab 3 - Excel Basics

Goals: Practice data handling, data cleaning, simple data analysis, visualization and prediction

Data: Population Estimates based on the 2010 US census

https://michael.hahsler.net/SMU/DS1300/data/census/sub-est2017_all.csv

Instructions

This lab will cover the first part of Assignment 3 - Where should our next campus be located?

(https://michael.hahsler.net/SMU/DS1300/material/Tables_Excel_Census.pdf)

- 1) Download the data and format it as a data table. [0]
- 2) Look at the dataset to understand what each column means and what columns are important. You want to perform analysis at the county level. Filter only county data (see SUMLEV) and create a new table with just this data [2 points]
- 3) Calculate population growth in each county (2010 to 2017) in number of people and also in percent. What county has the highest growth? [2 points]
- 4) Use a pivot table to aggregate the growth data to state level. What are the top 3 states in terms of population growth. [2 points]
- 5) Visualize population growth. Use a bar chart and a map chart. Discuss in a few sentences what you see. [2 points]
- 6) Use trendlines to try to predict the population development of Texas and California over the next 50 years. How do the two states compare? What growth model do you use and why? [2 points]

Submit the excel sheet via Canvas.