JOINT FLORIDA
Model Task Force & Transportation
Data and Analytics Workshop

FTO FORECASTING & TRENDS OFFICE
FDOT TDA TRANSPORTATION DATA & ANALYTICS

Performance Measures
Prediction of Future Performance

Dan Beaty & Dave Schmitt

Outline

- Introduction
- Where We Are
  - Use of Travel Demand Models
  - Types of Performance Measures
- Forward Thinking
  - Definitions
  - Issues from Explanatory Modeling / Data Interactions
  - Where Data and Models Successfully Work Together
- Discussion

Section 1 – Where We Are

Introduction

Issue: Prediction of Future Performance

- Current Performance Measures
  - Measure Past or at best Real Time Performance
  - Can be used for Trend Analysis
- Projecting Performance Measures
  - Can assisting in Development of Improvements
  - Can allow for Testing of Improvements before they are implemented

FDOT

Definitions

- Trend data/Time series models – forecasting based on the continuation of past/current trends
  - Example: average rate of traffic growth over past 5 years +10%

  → Traffic growth will be +10% in next 5 years
- Explanatory models – forecasting based on identified relationships
  - Example 1: Population growth +10%

  → Traffic growth should be +10%
  - Example 2: Population growth +10% and new Amazon warehouse

  → Traffic growth should be +400%

Use of Travel Demand Models

- Travel Demand Model Types
  - 4 Step Models
    - Daily Traffic
    - Peak Period or Peak Hour Traffic
  - Activity Based
    - Periods or down to 15 minutes increments of Traffic
    - Can test incidents, policies etc.

Types of Performance Measures

- Mobility
- Reliability
- Land Use/Urban Design
- Safety & Infrastructure
- Energy/Environment

Section 2 – Forward Thinking

Issues that Hinder Predictability of Future Performance

- Data is collected at different times using different definitions
  - Example 1: Job vs. job shift vs. employee
  - Example 2: Point speeds vs. average speed
- Trend depends on selected time frame:
  - Over past 5 years = average growth is +10%
  - Over past 10 years = average growth is -1%
  - So which is appropriate?

Issues that Hinder Predictability of Future Performance

- Selection bias, when the data outside our sample is not the same our collected sample
  - Example: assume LBS data is <5% sample…is the sample same or different from the other 95%? How do you know?
- More data → tendency for overfitting → challenges when forecasting
  - Example 1: separate volume / travel time relationships for each facility type
  - Example 2: separate volume / travel time relationships for each facility
  - New facility being constructed…which relationship to use?
- Data = Signal + Noise, but separating the two is challenging
  - Noise > signal in many cases
  - Spurious correlations (https://www.tylervigen.com/spurious-correlations)

FDOT

Spurious Correlations…

$r^2=0.66$

tylervigen.com

Source: https://www.tylervigen.com/spurious-correlations

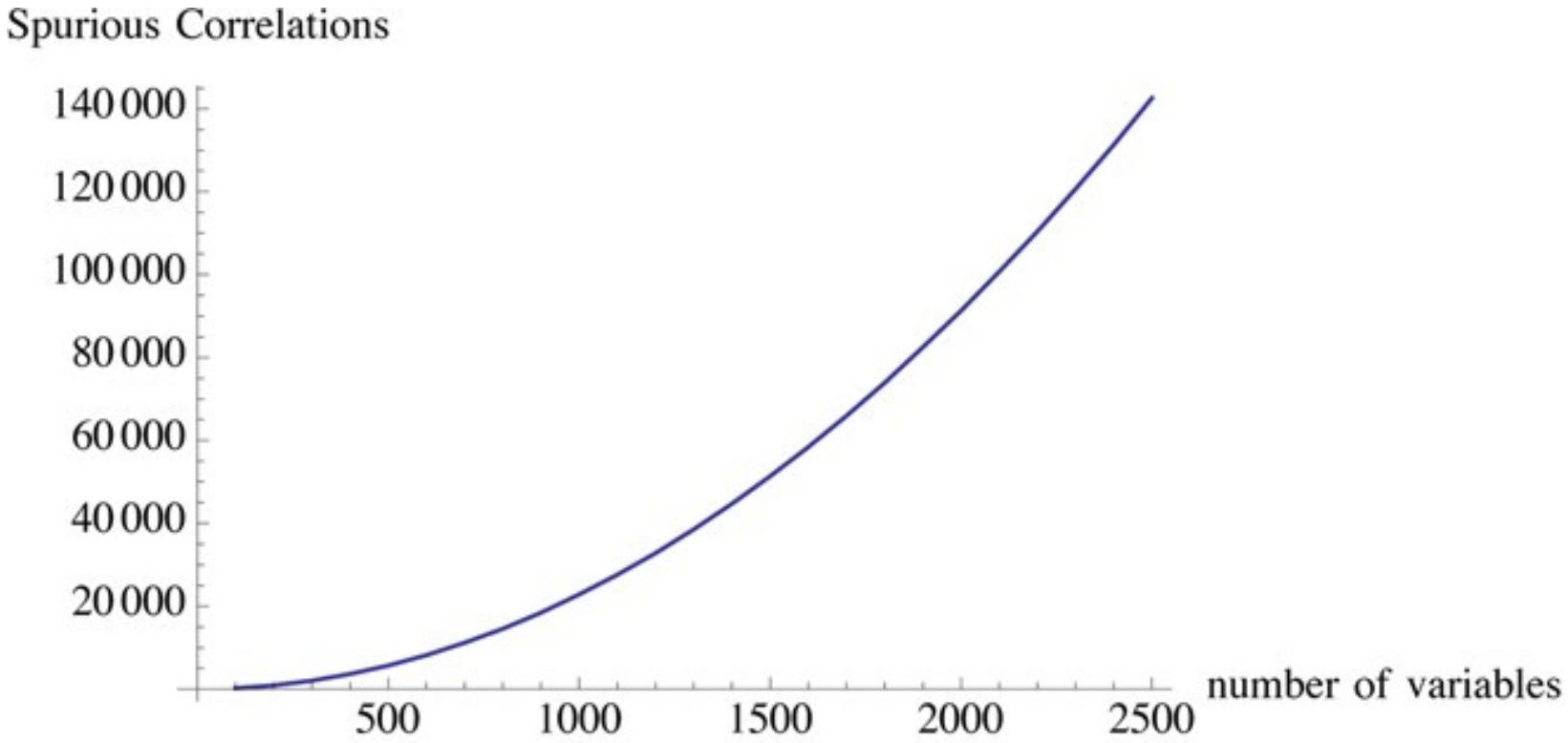
Another One…

$r^2=0.94$

tylervigen.com

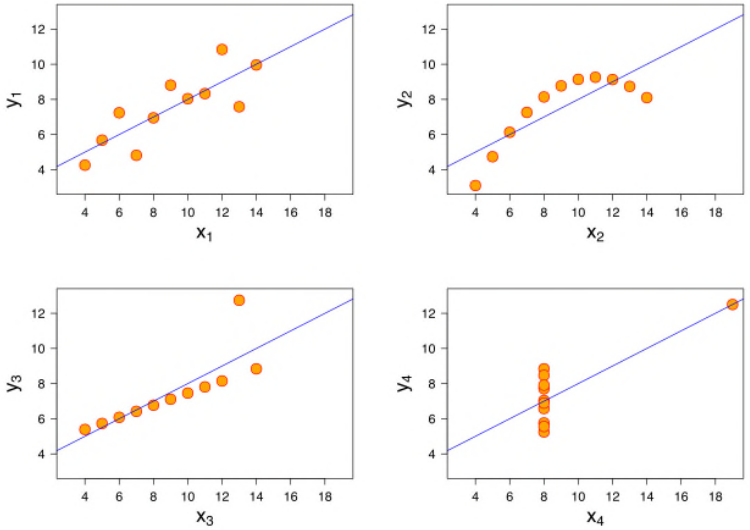
Source: https://www.tylervigen.com/spurious-correlations

On a Serious Note (1)



Source: NN Taleb, https://www.wired.com/2013/02/big-data-means-big-errors-people/

On a Serious Note (2)



**Four very different datasets,
but same regression line**

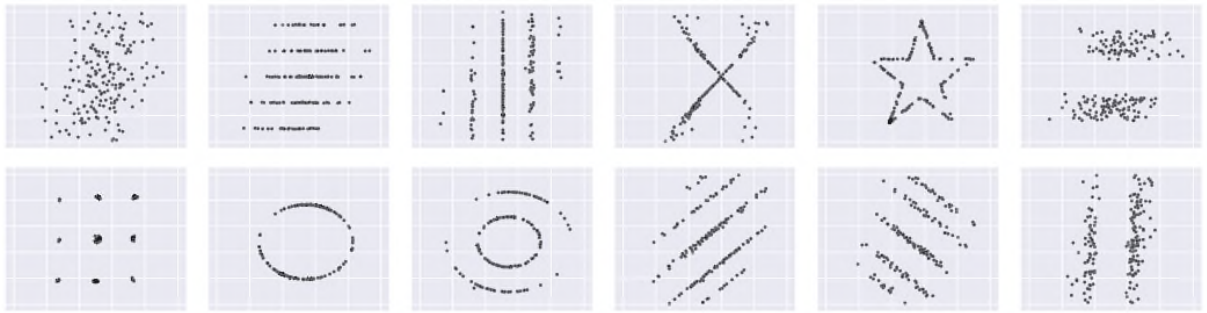


Figure 1. A collection of data sets produced by our technique. While different in appearance, each has the same summary statistics (mean, std. deviation, and Pearson's corr.) to 2 decimal places. ($\bar{x}$ *=54.02,* $\bar{y}$ *= 48.09,* sd_x *= 14.52,* sd_y *= 24.79, Pearson's r = +0.32)*

**Ten very different datasets,
but same mean,
standard deviation,
variance
and correlation…
…to 2 decimal places**

Sources:
http://footballphilosophy.org/encyclopedia/correlation/ (left)
"Same Stats, Different Graphs". Matejka and Fitzmuarice. Autodesk Research. (right)

Where Data and Models Successfully Work Together

- Prediction of Future Performance
- More diverse datasets + recently collected data → Better models
- Identifying relationships to better inform Travel Demand Models
  - Example: before/after impacts from new facility or traffic solution
  - How much did new facility improve mobility?
  - What contributed to improvements?

DISCUSSION