We Know What You Want
Predicting ILL Requests With ALIEN

Using Real-time Data, Statistics, and Predictive Analytics to Inform ILL and Collection Development Strategies

And so it begins…

We started with collecting large scale ILL data using OBILLSK

Online Based Inter-Library Loan Statistical Kit

Designed to provide consortial interlibrary loan statistics

Consists of three parts:

- Local client software
- SQL data analysis
- Visual presentation of turnaround metrics

Overview
Borrowing
Lending
Search

Relais Load Leveling
Institutional Comparisons
Calculation Methods

Scanning
Tracking

Upload File
Contact

Texas Tech University - Overview

Texas Tech University

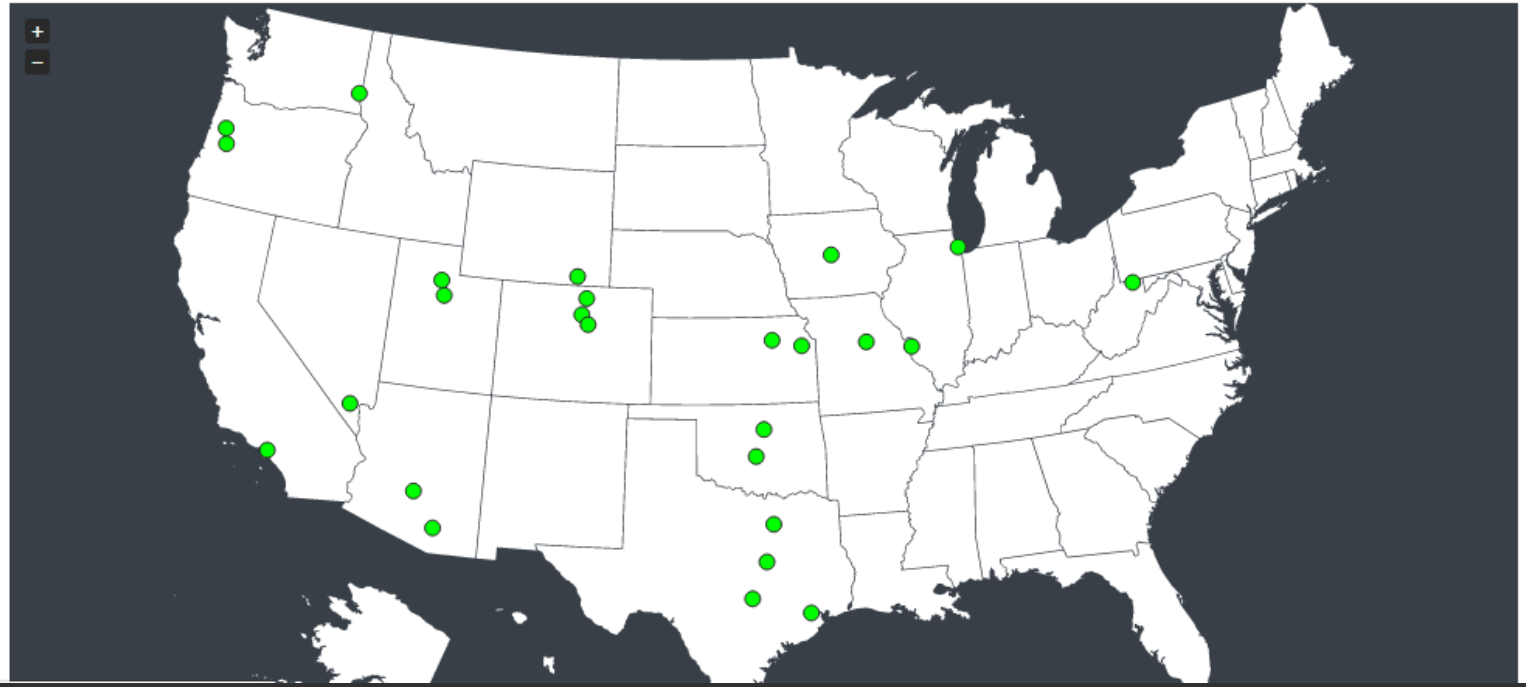
Institutional data last uploaded 9/13/2016 9:59:42 AM CT

Average Turnaround Times

Borrowing - Articles *0 Days, 4 Hours, 29 Minutes*	**In Transit - Articles** *0 Days, 4 Hours, 57 Minutes*	**Lending - Articles** *0 Days, 10 Hours, 24 Minutes*
Borrowing - Loans *0 Days, 4 Hours, 10 Minutes*	**In Transit - Loans** *1 Days, 18 Hours, 43 Minutes*	**Lending - Loans** *0 Days, 14 Hours, 2 Minutes*

Show All

Regional Transactions



What else can we do?

We realized the amount of data we were collecting allowed for new methods of analysis

The types of data collected were:

- Citation Information
- OCLC number

The data allows for a more in depth collection analysis using advanced statistical methods and predictive analytics

ALIEN knows…

ALIEN – the Automated Library Information Exchange Network

The beginnings of the first library machine learning decision network

A learning computer that examines a variety of library data sources to construct probabilistic and statistical decisions

Using predictive analytics with ILL, Reserves and Circulation data to recommend library behavior


I, FOR ONE,
WELCOME OUR NEW ROBOT OVERLORDS
memegenerator.net

How Predictive ILL Works

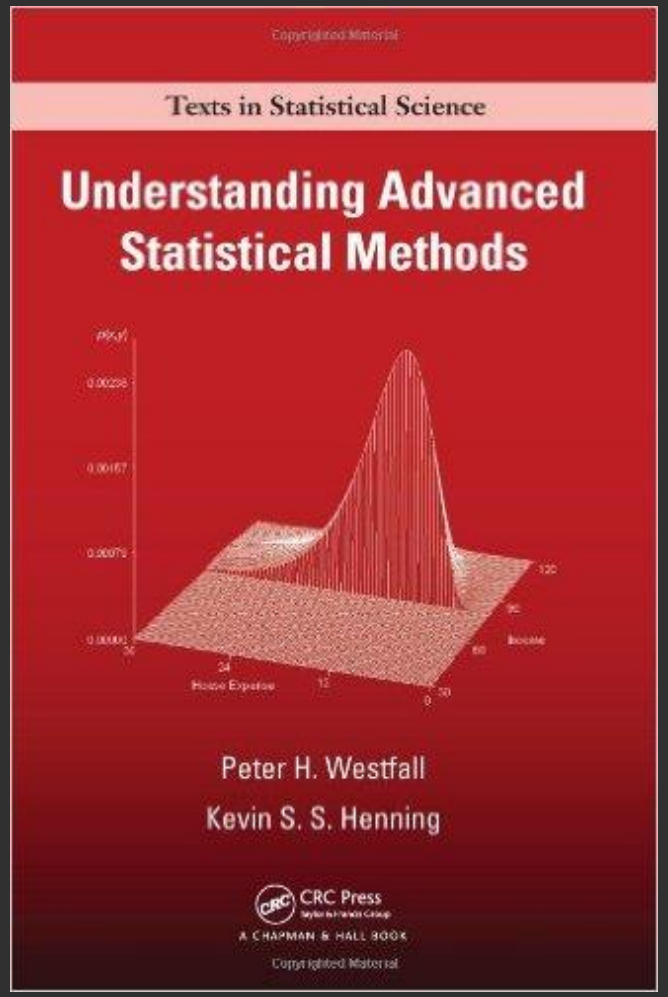
Using data gathered from the ILL servers, we create generations of ILL requests

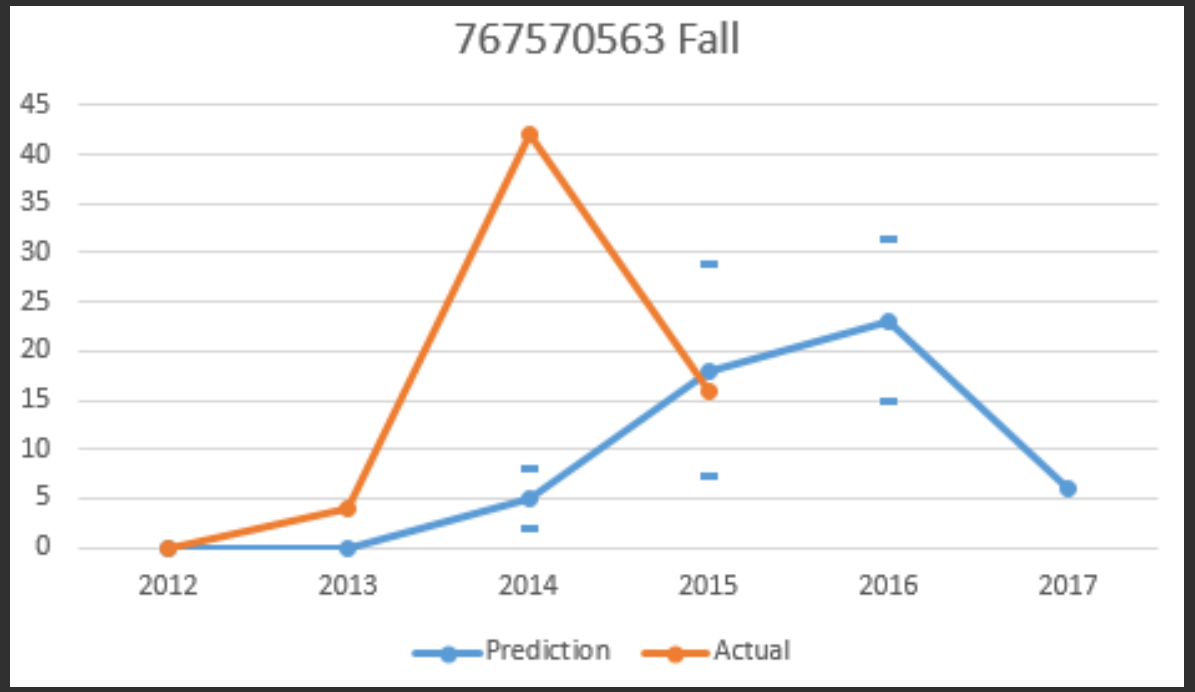
The system establishes a probabilistic curve for each semester (generation) for the ILL requests received for a certain OCLC number

Each successive generation, the system refines the curve and the confidence interval it gives for a prediction

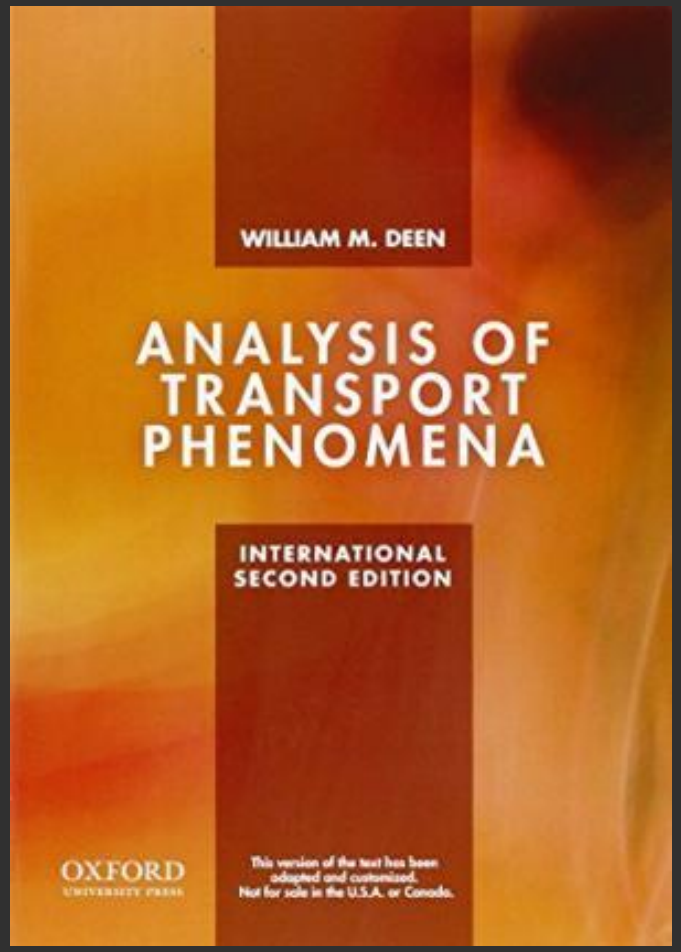
Eventually the curve will narrow to match reality

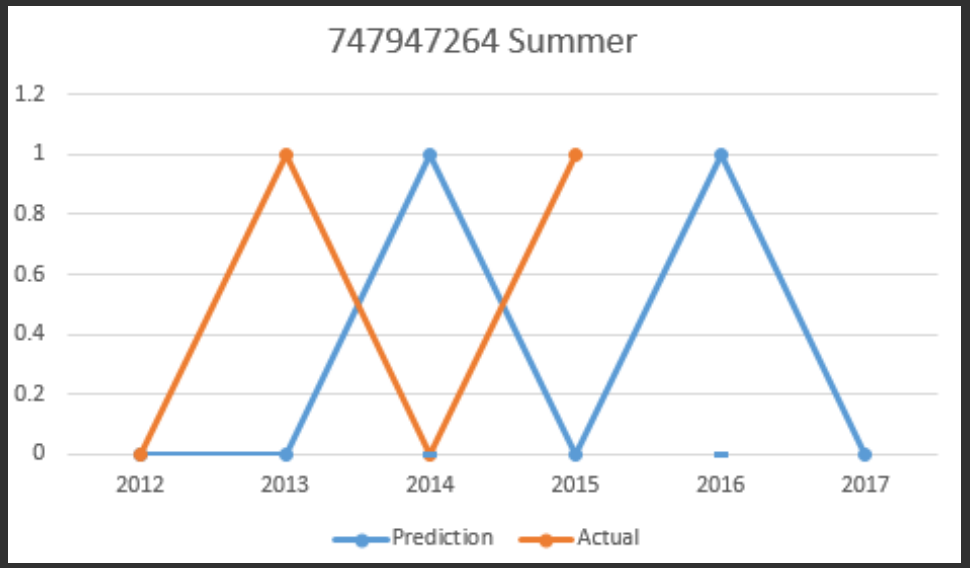
What Does It Look Like





What Does It Look Like





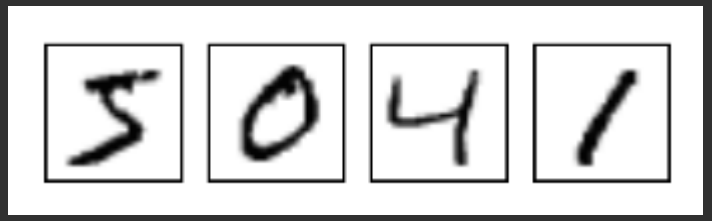
Moving Beyond ILL

ILL did not have enough training data

Using the ILL, Reserves and Circulation data and their OCLC numbers, we can match with WorldCat data to do our next level of analysis

Collection analysis is carried out with advanced statistical techniques including k-means clustering

Tensorflow



An open source software library for machine intelligence

k-Means

k-means clustering is a mathematical approach for large scale data analysis

It partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean

So we take n observations; ILL requests, circulation stats, in house use stats, course reserves

Each n observation for a specific data set (ILL, Reserves, Circ stats) is stored separately

We search OCLC using the OCLC number to determine the k cluster using WorldCat subject descriptors

An Example

Description

Publication:	New York, Hawthorn Books [1969]
Edition:	1st ed.].
Physical Description:	183 pages illustrations 22 cm
Language:	English
Staff View:	MARC Record
OCLC Number:	11932
LCCN:	69012958
Contents:	The vampire -- The family of a vourdalak -- The reunion after three hundred years -- Amena.
Subjects:	Vampires Fiction. Manners and customs. Vampires. Russia Social life and customs Fiction. Russia. Paranormal fiction. Short stories. Fiction.
Genre:	Paranormal fiction. Short stories. Fiction.

We use a combination of data sources to gather data on a libraries collection

We analyze each book for it's recognized OCLC subjects and genres

We sort each book into a subject and genre category

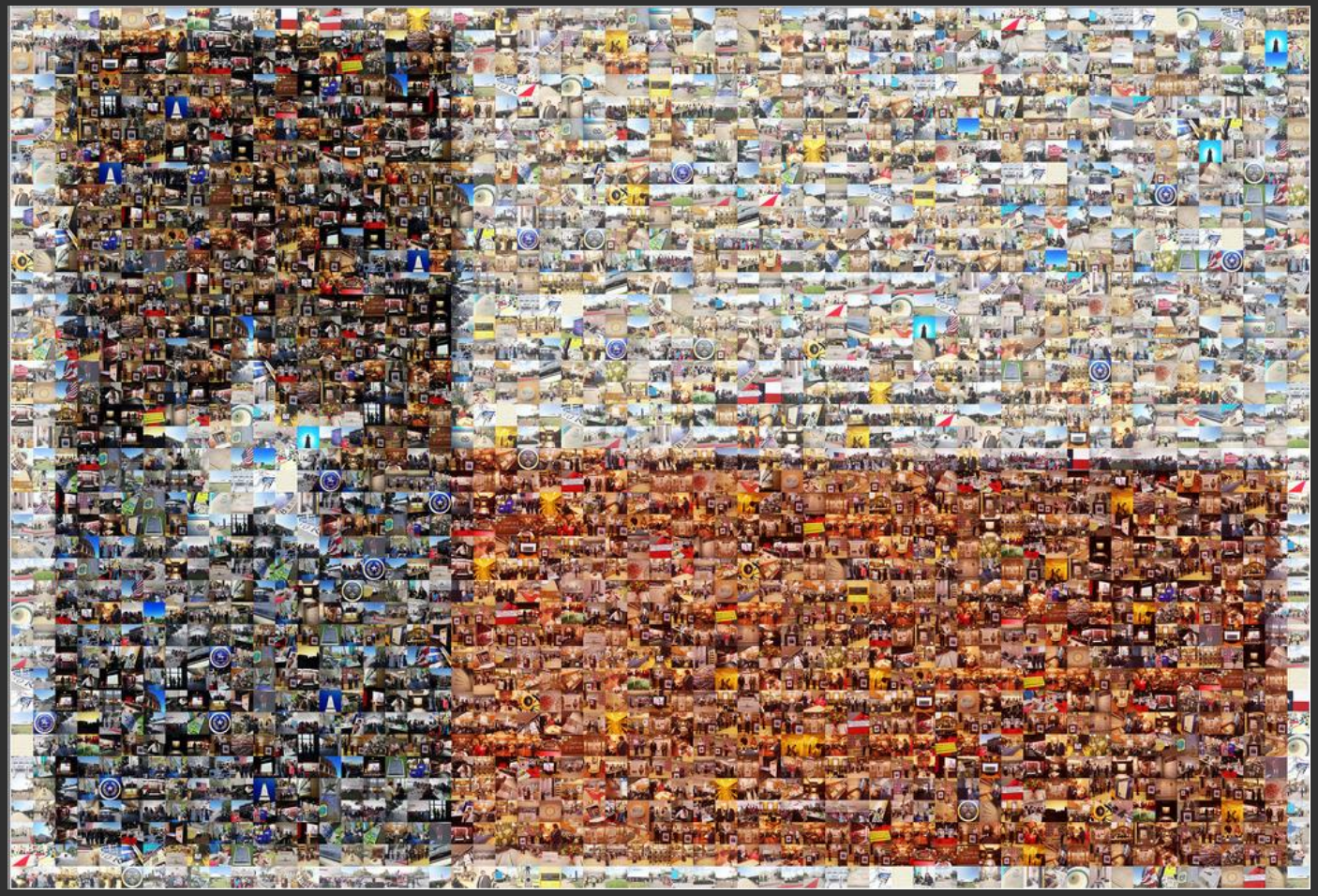
We analyze where the libraries collection has the most convergence – this represents your library's focus

What Does It Mean?

We are effectively using a variety of data sources to make large scale comparisons

We take one book and compare it to others using ILL requests, reserves, in-house use, circulation data, and place them in an overarching subject

Eventually a complete picture of the library emerges



Next Steps

For ALIEN we are looking to further develop:

- Accessing course reserves data
- Build recommendation functionality

Can we access facilities data for maintenance issues

If we can anticipate and predict library behavior does it change what we think of as assessment?

Add natural language UI

Thank You

Ryan Litsey

Ryan.Litsey@ttu.edu

Weston Mauldin

W.Mauldin@ttu.edu

Scott Luker

Scott.Luker@ttu.edu