# Data Analytics – An Industrial Engineering Perspective

 Fadel M. Megahed

 Assistant Prof. of Industrial & Systems Engineering

 Email: fmegahed@auburn.edu

 Web: www.fadelmegahed.com

Workshop Internacional de Ingeniería Aplicada Auburn - UACh Ian. 2015

COLLEGE OF ENGINEERING



## What is **Big Data**?

 Big data is a popular term used to describe the large, diverse, complex and/or longitudinal datasets generated from a variety of instruments, sensors and/or computer-based transactions.<sup>1</sup>

Google	"Big Data"										
	Web	Imágenes	Noticias	Videos	Libros	Más 🔻	Herramientas de búsqueda	búsqueda			
	Cerca de 62,400,000 resultados (0,23 segundos)										

<sup>1</sup> National Science Foundation. Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) <u>http://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=504767</u>. Last accessed 8/9/2013.

## What is **Big Data**?

 Big data is a popular term used to describe the large, diverse, complex and/or longitudinal datasets generated from a variety of instruments, sensors and/or computer-based transactions.<sup>1</sup>





<sup>1</sup> National Science Foundation. Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) <u>http://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=504767</u>. Last accessed 8/9/2013.

## What is **Big Data**?

 Big data is a popular term used to describe the large, diverse, complex and/or longitudinal datasets generated from a variety of instruments, sensors and/or computer-based transactions.<sup>1</sup>



<sup>1</sup> National Science Foundation. Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) <u>http://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=504767</u>. Last accessed 8/9/2013.

### What is Big Data? – The 3 Vs

 The term big data refers not only to the size or volume of data, but also to the variety of data and the velocity or speed of data accrual.<sup>1</sup>





<sup>1</sup> National Science Foundation. <u>http://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=504767</u>. Last accessed 8/9/2013. <sup>2</sup> Hyundai's Alabama Plant Image is from <u>http://www.businessalabama.com/Business-Alabama/March-2011/Hyundai-Could-Double-Down/</u>. Last accessed 8/9/2013.

## What is Big Data? – Beyond the 3 Vs

- The volume, variety and velocity of data distinguish big data from other types of data.
- Other V's important to **big data** problems are:
  - Veracity → Trustworthiness of the data
  - Value 
    → Added value to creating knowledge in a topic
- Both veracity and value relate to data quality

The Associated Press 💙 Follow The Fraudulent CAP tweet from the Breaking: Two Explosions in the White hacked AP account House and Barack Obama is injured Reply 13 Retweet \* Favorite ••• More An instant drop in 2.111 😹 🛃 🛐 🔍 🚎 🌆 🐴 🎆 🐖 the DOW Jones FAVORITES industrial average 11:07 AM - 23 Apr 13 A



Acquiring Data *≠* New Knowledge about System 9

"More information ought to be useful, but only if companies can interpret it. And workers are already overloaded: 62% of them say that the quality of what they do is hampered because they cannot make sense of the data they already have ..." The Economist

Deming (*Out of the Crisis, P.106*): "Information, no matter how complete and speedy, is not knowledge. Knowledge has temporal spread. Knowledge comes from theory. Without theory, there is no way to use the information that comes to us on the instant."



Data  $\rightarrow$  Information  $\rightarrow$  Knowledge

A

## Developing Knowledge or Data-based System 11 Models to Improve the Decision Making Process



### **Recent Research Sponsors**



Windows Azure



12

## **Application Domains**



## **Problem Description**

- A major US auto plant provided us with a dataset of 1000 vehicles, which had these characteristics\*:
  - 100% sampling was performed at 27 locations using CMMs
  - 39 measurements (deviations from nominal) per vehicle
- Objective: Assist in understanding process variation and identifying areas for quality improvement P13 (z) P13 (y, z) P25 (y, z) P25 (y, z) P27 (x, z)



\* Due to proprietary issues, the original data values, such as the number and location of sample points (black spheres in above figure), and the sample data have been slightly modified

## Our Framework for Visualization of Such Data

15

- Our method utilizes the strengths of PCA with the power of visualization tools (using VRML).
- It consists of : 1) Data preparation, 2) Data analysis and variation identification, and 3) Variation understanding through CAD visualization



### **Results and Significance – Immediate Results**



Wells, L.J., Megahed, F.M., Camelio, J.A., Woodall, W.H., 2012, "A Framework for Variation Visualization and Understanding in Complex Manufacturing Systems", *Journal of Intelligent Manufacturing*, 23 (5), pp. 2025-2036

## **Results and Significance – Extensions of Work**

#### Development of a Quality Visualization Toolkit (QVT) for Manufacturers – QVT Deployed in 2 AL Manufacturers



Smith, H.\*, Megahed, F. M., Jones-Famer, L.A., Clark, M. 2014, "Using Visual Data Mining to Enhance the Simple Tools in Statistical Process Control: A Case Study", *Quality and Reliability Engineering International*, to appear.

## **Results and Significance – Extensions<sup>2</sup>**

#### Student researchers spun out a company to assist local industry



18

## **Application Domains**



## **Understanding Traffic Patters for Trucking Safety** 20

Station :	000044	RI	I85 Alabama Department Of Transportation										Run Date: 3/30/2011 1:58:52 PM												
6.0 Miles	s South	of Macon Co. Line Transportation Planning Bureau - Traffic Division Direction Of Travel: North											orth												
Mt. Meigs Co: MONTGOMERY							Directional Monthly Volume Report									April 2010									
DAY	1 AM	2 AM	3 AM	4 AM	5 AM	6 AM	7 AM	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM	8 PM	9 P M	10 PM	11 PM	12 AM	TOTAL
1 T	296	199	173	167	299	425	728	1012	1003	1023	1150	1253	1353	1418	1533	1681	1881	1971	1446	1221	995	888	611	425	23151
2 F	268	207	192	201	254	356	589	837	1006	1146	1106	1216	1277	1412	1655	2088	2035	1776	1489	1217	1038	889	707	472	23433
3 S	366	204	191	182	169	233	369	587	743	856	1104	1143	1225	1141	1201	1217	1081	1064	1012	905	793	677	607	385	17455
4 S	255	168	121	109	111	174	223	380	552	712	965	1078	1258	1393	1507	1652	1788	1820	1758	1785	1272	841	609	400	20931
5 M	228	161	155	177	265	460	708	945	1082	1075	1207	1391	1316	1564	1473	1541	1829	1757	1152	994	743	647	410	292	21572
6 T	188	156	133	174	257	416	728	1060	1009	952	1085	1083	1275	1258	1317	1524	1569	1603	1266	911	798	682	445	287	20176
7 W	202	168	160	178	253	463	720	975	999	977	1085	1164	1344	1345	1539	1584	1830	1746	1236	901	735	638	434	324	21000
8 T	209	148	159	185	263	441	740	945	938	930	1075	1174	1359	1329	1560	1679	1844	1686	1272	1027	809	719	466	340	21297
9 F	218	186	153	193	269	437	785	1059	1015	1143	1222	1433	1623	1753	1850	2013	2139	2089	1736	1316	1201	1014	720	500	26067
10 S	311	206	159	172	208	355	466	754	901	1047	1283	1430	1593	1844	2040	1888	1627	1357	1199	1023	869	810	704	476	22722
11 S	333	186	157	139	153	188	304	496	769	1008	1250	1347	1764	1712	1658	2056	2054	1929	1434	1136	860	630	543	334	22440
12 M	223	162	122	165	249	421	730	911	953	946	998	1084	1138	1133	1213	1385	1591	1548	958	812	642	574	421	247	18626
13 T	161	124	118	166	242	429	763	978	902	864	953	927	1052	1102	1180	1298	1656	1572	1049	812	685	611	371	274	18289
14 W	209	131	149	154	248	439	749	1019	913	923	1011	1012	1133	1136	1313	1371	1614	1579	1103	821	722	584	401	287	19021
15 T	202	171	143	148	243	461	775	1064	916	1010	1167	1157	1208	1247	1314	1517	1759	1680	1212	971	799	750	432	291	20637
16 F	221	183	147	177	253	440	723	1041	1033	1045	1176	1418	1462	1558	1638	1760	2049	1864	1550	1266	1064	911	740	470	24189
17 S	268	202	177	195	226	316	477	787	1067	1608	1889	1591	1328	1136	1115	1235	1107	1095	1053	865	785	611	516	352	20001
18 S	264	174	184	157	124	164	263	418	586	757	982	1104	1262	1503	1448	1520	1466	1517	1146	1071	863	619	399	301	18292
19 M	196	143	128	138	254	448	727	950	894	879	965	1088	1097	1129	1149	1345	1597	1553	968	801	624	521	312	247	18143
20 T	158	127	134	158	249	442	754	946	933	873	864	874	1043	1102	1185	1305	1648	1543	956	808	664	587	367	267	17987
21 W	169	125	142	177	234	478	809	981	914	933	961	941	1140	1148	1260	1324	1540	1570	1042	863	684	573	393	316	18717
22 T	206	154	146	166	276	466	762	967	907	901	986	1035	1175	1184	1359	1422	1668	1677	1236	935	773	673	454	315	19843
23 F	221	164	145	189	271	415	769	995	990	1009	1083	1264	1369	1461	1619	1630	1859	1814	1441	1113	983	740	604	382	22530
24 S	252	193	144	162	171	278	371	529	645	757	965	912	993	953	947	952	865	813	701	619	487	469	416	309	13903
25 S	193	150	132	101	105	114	216	387	577	824	993	1164	1395	1463	1548	1639	1530	1536	1296	1133	863	611	489	325	18784
26 M	195	152	139	143	237	397	766	978	865	946	979	1044	1201	1174	1240	1340	1398	1465	987	798	644	546	327	273	18234
27 T	156	115	137	143	240	400	768	933	918	826	920	921	1115	1135	1169	1362	1559	1533	968	770	654	587	378	235	17942
28 W	203	139	125	170	250	441	810	986	950	937	862	938	1095	1132	1206	1240	1720	1578	1020	818	/19	613	384	289	18615
29 T	167	154	159	148	248	436	790	982	958	913	1011	1102	1086	1238	1276	1482	1667	1638	1193	913	817	641	449	320	19788
30 F	223	160	170	183	318	445	805	1022	1034	1028	1160	1223	1370	1537	1587	1801	1919	1859	1525	1199	1008	877	623	434	23510

L

## **Results and Significance – Immediate Results**

- 1) Development of a Quality Visualization Toolkit (QVT) QVT Deployed in 2 AL Manufacturers
- 2) Quantifying Seasonality through Data-Driven Methods for Inclusion in Statistical Models that can Assess the Impact of Public Policy on Truck Accidents.



Tsai, Y-T.\*, Smith, H.\*, Swartz, S.M., Megahed, F.M., 2014, "Using Visual Analytics to Enhance the Understanding of Occupational Safety Data", *Submitted for Publication*.

## **Application Domains**



22

### Acknowledgments

- This work was supported in part by Health Resources and Services Administration contract 234-2005-370011C. The content is the responsibility of the authors alone and does not necessarily reflect the views or policies of the Department of Health Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.
- Work is also in collaboration with a research group at UMass-Lowell.



## **Background and Motivation**

 A heart transplant is a surgery done on patients who have end-stage heart failure, where a diseased heart is replaced a healthy heart from a deceased donor.<sup>1</sup>

## • Statistics on Heart Transplantations:

- In the US, approximately 3,000 patients are on the waiting list on any given day.<sup>2</sup>
- Only, 2200 heart transplants occur annually with the following survival rates:<sup>1, 2</sup>
  - 88 % survive more than 1 year
  - 75 % survive more than 5 years
- 56 % survive for more than 10 years
   1. National Heart, Blood and Lung Institute (2012). What is a heart transplant? <u>http://www.nhlbi.nih.gov/health/health-topics/topics/ht/</u> 2. National Heart, Blood and Lung Institute (2012). What To Expect Before a Heart Transplant
  - http://www.nhlbi.nih.gov/health//dci/Diseases/ht/ht\_before.html

## **Research Objective and Contributions**

- Research Goal: Identify which variables contribute to the outcome of heart transplants through a data-based approach.
- Contributions:
  - Isolate the pre-operative variables affecting the short, medium and long-term survivability of heart transplant patients
  - Determine how these factors differ based on survival time
  - Is it possible to group these variables whose effect change over time ?

#### **The Dataset**

 A large, nationwide dataset obtained from United Network for Organ Sharing (UNOS), including ~43,000 patients and 437 variables.



## The Graphical Representation of the Methodology27



## Results and Significance – Straight from Models

		<u>AUC</u>	<u>Accuracy</u>	<u>Sensitivity</u>	<u>Specificity</u>
	CRT.S	$0.68 \pm 0.02$	$0.666 \pm 0.023$	$0.634 \pm 0.062$	$0.671 \pm 0.032$
al Its	CRT.U	$0.655 \pm 0.026$	$0.627 \pm 0.036$	$0.638 \pm 0.059$	$0.626 \pm 0.045$
viva	NN.S	$0.705 \pm 0.025$	$0.751 \pm 0.017$	$0.596 \pm 0.031$	$0.773 \pm 0.020$
ır Sur tion R	NN.U	$0.700 \pm 0.052$	$0.633 \pm 0.017$	$0.734 \pm 0.054$	$0.618 \pm 0.014$
	LR.S	$0.720 \pm 0.039$	$0.673 \pm 0.012$	$0.653 \pm 0.044$	$0.676 \pm 0.015$
_yea	LR.U	$0.701 \pm 0.032$	$0.655 \pm 0.009$	$0.622 \pm 0.041$	$0.660 \pm 0.011$
Pr J	SVM.S	$0.708 \pm 0.040$	$0.724 \pm 0.011$	$0.586 \pm 0.027$	$0.744 \pm 0.012$
	SVM.U	$0.699 \pm 0.028$	$0.592 \pm 0.010$	$0.680 \pm 0.046$	$0.580 \pm 0.011$
	CRT.S	$0.733 \pm 0.027$	$0.713 \pm 0.019$	$0.678 \pm 0.032$	$0.734 \pm 0.028$
al Its	CRT.U	$0.728 \pm 0.022$	$0.675 \pm 0.016$	$0.673 \pm 0.060$	$0.678 \pm 0.036$
viv	NN.S	$0.759 \pm 0.018$	$0.715 \pm 0.018$	$0.683 \pm 0.029$	$0.730 \pm 0.022$
Sur n R	NN.U	$0.731 \pm 0.026$	$0.649 \pm 0.020$	$0.688 \pm 0.050$	$0.628 \pm 0.023$
ar ctio	LR.S	$0.762 \pm 0.022$	$0.727 \pm 0.026$	$0.645 \pm 0.034$	$0.774 \pm 0.034$
_ye, edi	LR.U	$0.758 \pm 0.025$	$0.682 \pm 0.014$	$0.667 \pm 0.051$	$0.692 \pm 0.029$
Pr 5	SVM.S	$0.774 \pm 0.017$	$0.758 \pm 0.015$	$0.600 \pm 0.024$	$0.846 \pm 0.020$
	SVM.U	$0.753 \pm 0.027$	$0.664 \pm 0.019$	$0.702 \pm 0.049$	$0.643 \pm 0.034$
	CRT.S	$0.839 \pm 0.017$	$0.763 \pm 0.010$	$0.799 \pm 0.020$	$0.676 \pm 0.316$
al ilts	CRT.U	$0.821 \pm 0.021$	$0.741 \pm 0.018$	$0.680 \pm 0.028$	$0.881 \pm 0.025$
viv	NN.S	$0.845 \pm 0.020$	$0.778 \pm 0.021$	$0.831 \pm 0.015$	$0.653 \pm 0.045$
Sur n R	NN.U	$0.822 \pm 0.026$	$0.750 \pm 0.016$	$0.720 \pm 0.018$	$0.818 \pm 0.050$
itio	LR.S	$0.862 \pm 0.019$	$0.774 \pm 0.018$	$0.789 \pm 0.020$	$0.740 \pm 0.046$
yea	LR.U	$0.846 \pm 0.027$	$0.740 \pm 0.014$	$0.649 \pm 0.017$	$0.948 \pm 0.020$
Pre	SVM.S	$0.865 \pm 0.024$	$0.801 \pm 0.019$	$0.894 \pm 0.027$	$0.600 \pm 0.015$
	SVM.U	$0.856 \pm 0.024$	$0.752 \pm 0.022$	$0.713 \pm 0.022$	$0.843 \pm 0.047$

L

### **Results and Significance – What do they Mean?**

29



Dag, A.\*, Megahed, F. M., Oztekin, A., Chen, Y., Yucel, A., 2014, "Identifying Predictor Variables for the Success of Heart Transplants", *European Journal of Operations Research* (Under review).

## **Application Domains**



30

#### **Objectives**

Marketing Distribution Warehousing Optimize Stocking Predict vs. Respond Increase sales Reduce Costs

#### **Data Sources**

2

Twitter Facebook Google HealthMap FluNearYou CDC/WHO Sales/Inventory Data

#### How?

3

Optimization Clustering Surveillance Data management Other methods

#### **Previous Work**

Applied analytics to many industries Healthcare logistics, distribution Social media data mining

2

#### 5 Our Network

Auburn University UCF Miami at Ohio Auburn Pharmacy Auburn Software

## **Data Sources**

- Twitter
- Google Trends & Plus

A

- Facebook
- Wikipedia
- FluNearYou
- CDC/WHO
- AthenaHealth
- Others



## **Objectives/Ideas**

- Apply tool to marketing retweet based on location
  - Increased sales
- Distribution/Warehousing
  - Optimize locations & timing
- Optimization of Stocking Levels
  - Minimize under/over stocking
- Predict vs. Respond
  - Based on aggregate data







## **Data Sources – Cross Validation**

-2

## Having multiple data sources work in tandem is necessary.





## **Data Analytics Opportunities for IEs in the Past** 35





## **Current and Future Opportunities**





# Data Analytics – An Industrial Engineering Perspective

 Fadel M. Megahed

 Assistant Prof. of Industrial & Systems Engineering

 Email: fmegahed@auburn.edu

 Web: www.fadelmegahed.com

Workshop Internacional de Ingeniería Aplicada Auburn - UACh Ian. 2015

COLLEGE OF ENGINEERING