

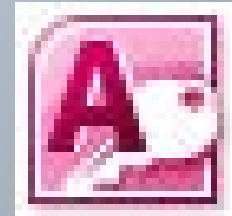
Data Mining: Using e-book data to visualize and analyze a collection

Yvonne Lee, MS, Laura Abate, MSLS, Julie Silverman, MLS, MBA
Himmelfarb Health Sciences Library, George Washington University

Methods

Source data compiled in MS Access

- Serialsolutions report with broad subject categories for department assignments
- E-Book usage data
- Faculty counts by department
- Data cleaned up & merged by ISBN/E-ISBN
- Exported merged data into Excel



Visualization Charts & Graphs

Data analyzed and displayed in Excel

- Excel PivotTables/Charts with Slicers filter Schools and Departments
- Interactive dashboard with hyperlinks increases flexibility and ease of navigation
- Designed to simulate a webpage experience
- Detailed title lists allow staff to move between charts/graphs and underlying title lists



Results

Analysis examples

- Size of collection and usage by Department & School
- Relationship of faculty size, collection size and usage
- Currency of collection by title count by publication years
- Usage comparison by consecutive years by Department
- E-book usage by Vendor by School

Conclusion

Effective data mining and visualization

- Provides broad and detailed collection data to support decision making
- Creates innovative ways of viewing and analyzing e-book collection
- Identifies areas which need additional data
- Permits comparison of Departments and Schools
- Promotes efficient fund management and better collection support for Schools and Departments
- Demonstrates Excel's powerful tools and advanced features

Excel Hyperlink

2015 E-Books: Title Count by Publication Year by Department

Year	Count
2003	1
2010	2
2011	2
2012	4
2013	4
2014	6
2015	1
2016	2

2015 E-Book Total Usage by Dept. with Title Count

Dept	Usage	Title Count
Med-Gen Internal Medicine	5768	2682
Nursing	318	170
Public Health	170	45

SMHS Faculty & Dept Statistics

Department	Faculty	No. of Titles	Total Usage	Titles per Faculty	Usage per Faculty
Med-Genomic Medicine	2	23	4301	0.61%	2150.50
CRL-Clinical Mgmt/Leadership	8	129	10458	3.42%	1307.25
CRL-Clinical Res Admin	4	55	4631	1.46%	1157.75
Med-General Internal Medicine	42	126	46618	3.34%	1109.95
Dermatology	5	70	4135	1.85%	827.00
CRL-Clinical/Translational Res	1	6	508	0.16%	508.00
Anatomy & Regenerative Biol	20	78	10064	2.07%	503.20
Pharmacology & Physiol	21	124	9757	3.28%	464.62
Emergency Medicine	41	102	17082	2.70%	416.63
Med-Pulmonary/Crit Care/Sleep Dis	8	54	2917	1.43%	364.63
Obstetrics & Gynecology	28	138	9915	3.65%	354.11
Pathology	28	175	9862	4.63%	352.21
Orthopedic Surgery	19	221	5926	5.85%	311.89
Medical Toxicology-GME	1	12	273	0.32%	273.00
MED-Genetics & Palliative Care	18	51	4639	1.35%	257.72
Otolaryngology-GME	10	20	2436	0.53%	243.60
Nursing	59	483	13249	12.79%	224.56
Surgery	60	266	12763	7.04%	212.72
Pediatrics	152	218	32216	5.77%	211.99
Med-Gastroenterol & Liver Dis	11	49	2102	1.36%	191.09
Med-Hematology & Oncology	12	135	2265	3.57%	188.75
Med-Cardiology	23	145	3915	3.84%	170.22
CRL-Medical Laboratory Sciences	1	8	150	0.21%	150.00
Physical Therapy	13	30	1930	0.79%	148.46
Med-Infectious Diseases	14	50	2013	1.32%	143.79
Microbiol Immunol Tropical Med	15	43	2153	1.45%	143.53
Neurology	48	167	6112	4.42%	127.33
CRL-Occupational Therapy	2	14	249	0.37%	124.50

Himmelfarb Library E-Book Analysis with Data Visualization

Department Titles	2014	2015	Grand Total
CURRENT Medical Diagnosis and Treatment	5993	11341	17334
2015	5993		5993
2016		11341	11341
Symptom to Diagnosis: An Evidence-Based Guide 3e	868	7021	7889
2014	868		868
2014		7021	7021
Harrison's Principles of Internal Medicine 19e		4968	4968
2015		4968	4968
Harrison's Principles of Internal Medicine 18e	6497	4513	11010
2012	6497		6497
2012		4513	4513
Textbook of family medicine	917	4217	5134
2015	917		917
2016		4217	4217
Ferri's clinical advisor	2989	1719	4708
2015	2989		2989
2016		1719	1719
Patient History	1060	1472	2532
2012	1060		1060
2012		1472	1472

E-Book: Top 15 Providers with Highest Usage

Provider	Usage
AccessMedicine	~85000
ClinicalKey	~75000
ClinicalKey for Nursing	~25000
R2 Library	~30000
Books@Ovid: LWW Doody's	~15000
AccessSurgery	~10000
McGraw-Hill	~5000
EBSCOhost	~5000
AccessAnesthesiology	~5000