

# Final Project

```
ccl <- rbind(ccl201601,ccl201602,ccl201603,ccl201604,ccl201605,ccl201606,ccl201607,ccl201608,ccl201609,
spu <- rbind(spu201601,spu201602,spu201603,spu201604,spu201605,spu201606,spu201607,spu201608,spu201609,
combined1 <- left_join(ccl,spu,by="session")
combined1 <- left_join(combined1,patrons,by="uuid")
combined1$campus.x[combined1$campus.x=="pom"] <- c("pomona")
combined1$campus.x[combined1$campus.x=="scr"] <- c("scrippscollege")
combined1$campus.x[combined1$campus.x=="pit"] <- c("pitzer")
combined1$domain <- as.factor(combined1$domain)
separate(combined1,datetime,c("year","month","day"),by=" ")
```

```
## Warning: Using ... for passing arguments to `strsplit()` is defunct
```

```
## Warning: Too many values at 446002 locations: 1, 2, 3, 4, 5, 6, 7, 8, 9,
## 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...
```

```
## Warning: Too few values at 8435 locations: 237646, 237647, 237648, 237649,
## 237650, 237651, 237652, 237653, 237654, 237655, 237656, 237657, 237658,
## 237659, 237660, 237661, 237662, 237663, 237664, 237665, ...
```

```
## # A tibble: 454,437 x 16
```

```
##                                     uuid
## *                                <chr>
## 1 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 2 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 3 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 4 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 5 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 6 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 7 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 8 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 9 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## 10 1493f5de47c3bcb37ed34a8c65e4cb85da71da2278126a29114b7e77bd827200
## # ... with 454,427 more rows, and 15 more variables: campus.x <chr>,
## #   year <chr>, month <chr>, day <chr>, session <chr>, ipSubnet <chr>,
## #   scheme <chr>, subdomain <chr>, domain <fctr>, tld <chr>, path <chr>,
## #   query <chr>, fragment <chr>, campus.y <chr>, patronType <chr>
```

```
combined1 <- combined1 %>%
```

```
  mutate(dateIN=substr(datetime,1,11),TimeIn=substr(datetime,12,20),TimeOut=substr(datetime,21,45),Month=
```

```
combined1$MonthUse[combined1$MonthUse==c("01")] <- c("01Jan")
combined1$MonthUse[combined1$MonthUse==c("02")] <- c("02Feb")
combined1$MonthUse[combined1$MonthUse==c("03")] <- c("03Mar")
combined1$MonthUse[combined1$MonthUse==c("04")] <- c("04Apr")
combined1$MonthUse[combined1$MonthUse==c("05")] <- c("05May")
combined1$MonthUse[combined1$MonthUse==c("06")] <- c("06June")
combined1$MonthUse[combined1$MonthUse==c("07")] <- c("07July")
combined1$MonthUse[combined1$MonthUse==c("08")] <- c("08Aug")
combined1$MonthUse[combined1$MonthUse==c("09")] <- c("09Sept")
combined1$MonthUse[combined1$MonthUse==c("10")] <- c("10Oct")
combined1$MonthUse[combined1$MonthUse==c("11")] <- c("11Nov")
combined1$MonthUse[combined1$MonthUse==c("12")] <- c("12Dec")
```

*#Unique Users - Trying to Find the Number of Distinct Users*

```
unique1 <- combined1 %>%
  filter(!is.na(patronType),!is.na(domain),!is.na(uuid)) %>%
  group_by(domain) %>%
  summarise(UniqueUsers=n_distinct(uuid)) %>%
  arrange(desc(UniqueUsers))

unique1 <- data.frame(unique1)
unique1
```

##	domain	UniqueUsers
## 1	ebscohost	6566
## 2	oclc	4907
## 3	ebrary	4189
## 4	jstor	4107
## 5	proquest	4022
## 6	doi	3199
## 7	sciencedirect	2875
## 8	tandfonline	2700
## 9	wiley	2578
## 10	sagepub	2476
## 11	springer	2193
## 12	ebilib	2106
## 13	galegroup	1651
## 14	isiknowledge	1489
## 15	lexisnexis	1293
## 16	jhu	1131
## 17	oxfordjournals	1016
## 18	cambridge	985
## 19	google	928
## 20	heinonline	725
## 21	claremont	706
## 22	handle	513
## 23	uchicago	440
## 24	nature	438
## 25	worldcat	438
## 26	webofknowledge	437
## 27	ieee	388
## 28	bioone	325
## 29	chadwyck	321
## 30	oed	302
## 31	apa	292
## 32	cro3	285
## 33	newsbank	247
## 34	mergentonline	243
## 35	oxfordscholarship	232
## 36	alexanderstreet	217
## 37	marketresearch	207
## 38	umich	206
## 39	aspresolver	205
## 40	springerlink	198
## 41	chicagomanualofstyle	194
## 42	naxosmusiclibrary	180

## 43	refworks	173
## 44	lwwhealthlibrary	163
## 45	cas	150
## 46	mhmedical	150
## 47	annualreviews	149
## 48	oxfordmusiconline	149
## 49	timesonline	148
## 50	lexi	147
## 51	els-cdn	146
## 52	pharmacylibrary	145
## 53	acs	138
## 54	accesspharmacy	132
## 55	oxfordreference	132
## 56	micromedexsolutions	131
## 57	dukejournals	128
## 58	eastview	120
## 59	ebsco	118
## 60	ingenta	117
## 61	worldbank	116
## 62	cqpress	115
## 63	emeraldinsight	115
## 64	ovid	115
## 65	llmcdigital	110
## 66	brillonline	105
## 67	acm	102
## 68	nejm	101
## 69	pdcnet	100
## 70	informs	97
## 71	kanopystreaming	96
## 72	uptodate	95
## 73	amdigital	94
## 74	credoreference	92
## 75	jamanetwork	91
## 76	ciaonet	90
## 77	nybooks	90
## 78	cnki	89
## 79	kstudy	88
## 80	westlaw	85
## 81	globalfinancialdata	83
## 82	imf	82
## 83	ams	81
## 84	factsandcomparisons	80
## 85	ucla	80
## 86	ucpress	80
## 87	abc-clio	76
## 88	artstor	76
## 89	metapress	74
## 90	philpapers	70
## 91	loebclassics	69
## 92	oxfordhandbooks	69
## 93	uspnf	68
## 94	ingentaconnect	66
## 95	chronicle	65
## 96	mitpressjournals	63

## 97	mlajournals	63
## 98	therapeuticresearch	58
## 99	degruyter	57
## 100	nlx	57
## 101	upenn	57
## 102	asm	56
## 103	oecd-ilibrary	56
## 104	tcrecord	56
## 105	crcnetbase	55
## 106	hoovers	52
## 107	aeaweb	51
## 108	newleftreview	51
## 109	referenceusa	51
## 110	oxfordartonline	50
## 111	pnas	50
## 112	mergent	49
## 113	prsgroup	49
## 114	aps	48
## 115	data-planet	48
## 116	geoscienceworld	48
## 117	accessible	47
## 118	conquestsystems	47
## 119	eb	47
## 120	nber	44
## 121	metopera	42
## 122	aip	41
## 123	medicinescomplete	40
## 124	sciencemag	40
## 125	browzine	39
## 126	rsc	39
## 127	amsciepub	37
## 128	humankinetics	37
## 129	nih	37
## 130	iop	36
## 131	psychiatryonline	36
## 132	jneurosci	35
## 133	oxforddnb	35
## 134	routledge	35
## 135	serialssolutions	35
## 136	valueline	35
## 137	yale	35
## 138	epw	33
## 139	anthrosource	32
## 140	netlibrary	31
## 141	ap	30
## 142	edweek	30
## 143	ipasource	30
## 144	swankmp	30
## 145	biologists	29
## 146	jbc	29
## 147	nationaljournal	29
## 148	virginia	29
## 149	escj	28
## 150	newstatesman	27

## 151	uconn	27
## 152	statista	26
## 153	chinadataonline	25
## 154	naxosvideolibrary	25
## 155	physiology	25
## 156	projecteuclid	25
## 157	currenthistory	24
## 158	oxforddictionaries	24
## 159	familiesinsociety	23
## 160	ulrichsweb	23
## 161	universitypublishingonline	23
## 162	brepolis	22
## 163	simplymap	22
## 164	ncte	21
## 165	northeastern	21
## 166	ripmfulltext	21
## 167	dyabola	20
## 168	palgraveconnect	20
## 169	anb	19
## 170	dramonline	19
## 171	merip	19
## 172	uci	19
## 173	uclajournals	19
## 174	routledgeperformancearchive	17
## 175	spectator	17
## 176	maneyonline	16
## 177	physicstoday	16
## 178	csa	15
## 179	elsevier	15
## 180	latinnews	15
## 181	liebertpub	15
## 182	clarku	14
## 183	pomona	14
## 184	timeshighereducation	14
## 185	esajournals	13
## 186	europaworld	13
## 187	hepg	13
## 188	itergateway	13
## 189	palgrave-journals	13
## 190	readcube	13
## 191	chosun	12
## 192	e-enlightenment	12
## 193	ebsco-content	12
## 194	osapublishing	12
## 195	oxfordaasc	12
## 196	royalsocietypublishing	12
## 197	telospress	12
## 198	cornell	11
## 199	harpweek	11
## 200	japantimes	11
## 201	lexis-nexis	11
## 202	newscientist	11
## 203	siam	11
## 204	thieme-connect	11

## 205	antiquity	10
## 206	elgaronline	10
## 207	fasebj	10
## 208	morganclaypool	10
## 209	standardandpoors	10
## 210	worldshakesbib	10
## 211	amjbot	9
## 212	igi-global	9
## 213	rochester	9
## 214	the-tls	9
## 215	aasa	8
## 216	ajol	8
## 217	americanscientist	8
## 218	archivegrid	8
## 219	aslo	8
## 220	hepgjournals	8
## 221	illinois	8
## 222	int-res	8
## 223	marquiswhoswho	8
## 224	aisnet	7
## 225	allenpress	7
## 226	capitaliq	7
## 227	choicereviews	7
## 228	columbia	7
## 229	dictionaryofeconomics	7
## 230	donga	7
## 231	ed	7
## 232	latinamericannewsdigest	7
## 233	medievalsources	7
## 234	nikkei	7
## 235	oriprobe	7
## 236	acrl	6
## 237	ajsonline	6
## 238	crl	6
## 239	globalsecurity	6
## 240	iiss	6
## 241	oecd	6
## 242	oxfordwesternmusic	6
## 243	portlandpress	6
## 244	rupress	6
## 245	annee-philologique	5
## 246	armand-colin	5
## 247	aspetjournals	5
## 248	blackwell-synergy	5
## 249	burlington	5
## 250	cdlib	5
## 251	cell	5
## 252	colmex	5
## 253	commentarymagazine	5
## 254	databanksinternational	5
## 255	dukeupress	5
## 256	duxiu	5
## 257	embopress	5
## 258	eupublishing	5

## 259	harpers	5
## 260	hbw	5
## 261	muohio	5
## 262	neurology	5
## 263	pitt	5
## 264	plantphysiol	5
## 265	rcpsych	5
## 266	rsnz	5
## 267	umi	5
## 268	aacrjournals	4
## 269	ala	4
## 270	aluka	4
## 271	ametsoc	4
## 272	ashspublications	4
## 273	biomedcentral	4
## 274	cshlp	4
## 275	grovemusic	4
## 276	indiastat	4
## 277	jaah	4
## 278	nagt	4
## 279	socialexplorer	4
## 280	unb	4
## 281	unige	4
## 282	worldscientific	4
## 283	worldscinetarchives	4
## 284	arabworldgeographer	3
## 285	biolbull	3
## 286	chemistrycentral	3
## 287	choice360	3
## 288	csiro	3
## 289	ele-math	3
## 290	emath	3
## 291	emis	3
## 292	harvard	3
## 293	hbcnetbase	3
## 294	ifar	3
## 295	jollas	3
## 296	karger	3
## 297	molecule	3
## 298	msp	3
## 299	myendnoteweb	3
## 300	operanews	3
## 301	princeton	3
## 302	scopus	3
## 303	thomsonreuters	3
## 304	ualberta	3
## 305	un	3
## 306	uwpress	3
## 307	vanderbilt	3
## 308	youtube	3
## 309	alliance1	2
## 310	arabidopsis	2
## 311	berghahnbooks	2
## 312	chbeck	2

## 313	co-psychiatry	2
## 314	columbiagazetteer	2
## 315	ems-ph	2
## 316	endocrine	2
## 317	endojournals	2
## 318	escholarship	2
## 319	eupjournals	2
## 320	familiesinsocietyjournal	2
## 321	genetics	2
## 322	impan	2
## 323	informahealthcare	2
## 324	japanlinkcenter	2
## 325	loc	2
## 326	math	2
## 327	nikkeiamerica	2
## 328	oopen	2
## 329	reference-global	2
## 330	reviews	2
## 331	sauerlaender-verlag	2
## 332	sif	2
## 333	stanford	2
## 334	theatlantic	2
## 335	washingtonpost	2
## 336	66.170.18.227	1
## 337	aanda	1
## 338	aanzpa	1
## 339	about	1
## 340	advancesindifferenceequations	1
## 341	aeon	1
## 342	agrifoodecon	1
## 343	agu	1
## 344	aiaa	1
## 345	akademai	1
## 346	ama-assn	1
## 347	amb-express	1
## 348	american-music	1
## 349	ammonsscientific	1
## 350	amses-journal	1
## 351	annalsofintensivecare	1
## 352	aom	1
## 353	apimages	1
## 354	applied-informatics-j	1
## 355	appliedadhesionscience	1
## 356	apsnet	1
## 357	archive	1
## 358	artinprint	1
## 359	arxiv	1
## 360	ascelibrary	1
## 361	ascimaging	1
## 362	astm	1
## 363	benjamins	1
## 364	biochemsoctrans	1
## 365	bit	1
## 366	boingboing	1



## 367	botany	1
## 368	businessinsider	1
## 369	catalystjournal	1
## 370	cedram	1
## 371	cfapubs	1
## 372	cgu	1
## 373	chemeducator	1
## 374	chemistry	1
## 375	computingreviews	1
## 376	copyright	1
## 377	crcpress	1
## 378	dailynexus	1
## 379	dankaminsky	1
## 380	dblp	1
## 381	earlytheatre	1
## 382	earthtimes	1
## 383	editlib	1
## 384	electrochem	1
## 385	epnet	1
## 386	ereaderiq	1
## 387	esa	1
## 388	europemc	1
## 389	feministstudies	1
## 390	galenet	1
## 391	geolsoc	1
## 392	getty	1
## 393	gging	1
## 394	greyhouse	1
## 395	gutenberg	1
## 396	hathitrust	1
## 397	hbr	1
## 398	headtraumarehab	1
## 399	hindawi	1
## 400	historyearthscience	1
## 401	hkusa	1
## 402	hljournals	1
## 403	hwwilsonweb	1
## 404	imdb	1
## 405	informaworld	1
## 406	informit	1
## 407	inverse	1
## 408	iopscience	1
## 409	ircps	1
## 410	jimmunol	1
## 411	jlir	1
## 412	jst	1
## 413	libraryjournal	1
## 414	llmc	1
## 415	lwbooks	1
## 416	lyellcollection	1
## 417	makezine	1
## 418	mendeley	1
## 419	metoperafamily	1
## 420	mgg-online	1

## 421	mineralogicalassociation	1
## 422	mobilizationjournal	1
## 423	monash	1
## 424	mtsac	1
## 425	nace	1
## 426	nationalgeographic	1
## 427	neurotodayonline	1
## 428	nomos-elibrary	1
## 429	numdam	1
## 430	nursingresearchonline	1
## 431	nytimes	1
## 432	ohiostatepress	1
## 433	opticsexpress	1
## 434	oup	1
## 435	oxfordlanguagedictionaries	1
## 436	pcworld	1
## 437	peeters-leuven	1
## 438	perspectivesofnewmusic	1
## 439	phycologia	1
## 440	plos	1
## 441	psmedia	1
## 442	psychiatrist	1
## 443	rcni	1
## 444	redalyc	1
## 445	repec	1
## 446	reviverestore	1
## 447	rism	1
## 448	rlg	1
## 449	scielo	1
## 450	scienceheathen	1
## 451	sciencesocieties	1
## 452	searchproquest	1
## 453	sfaajournals	1
## 454	sierraclub	1
## 455	socio	1
## 456	steiner-verlag	1
## 457	stonybrook	1
## 458	swetswise	1
## 459	synonym	1
## 460	tandfebooks	1
## 461	tcetoday	1
## 462	tcj	1
## 463	ted	1
## 464	terrapub	1
## 465	theconversation	1
## 466	thirdiron	1
## 467	thomist	1
## 468	tinygaleurl	1
## 469	truman	1
## 470	uillinois	1
## 471	uncg	1
## 472	uniromal	1
## 473	unl	1
## 474	uva	1

```

## 475          uwe          1
## 476        warwick        1
## 477      wikipedia        1
## 478      womenhiv         1
## 479      wordpress        1

# Journals that have less than 10 unique users

# 144 Journals have only 1 unique user. 269 Journals have less than 10 unique users.

# The average person associated with the Claremont Colleges uses 7.17 journals. The median journal usage is 5 journals.

# For Pomona -
usagepom <- combined1 %>%
  filter(!is.na(campus.x), !is.na(uuid), !is.na(domain), campus.x=="pomona") %>%
  group_by(uuid) %>%
  summarise(numofjournalsused=n_distinct(domain)) %>%
  summarise(avg = mean(numofjournalsused), median=median(numofjournalsused))

usagepom

## # A tibble: 1 x 2
##       avg median
##   <dbl> <int>
## 1 6.65098     5

# The average Pomona student uses 6 different journals, while the median use is 7 journals.

# CMC usage
usagecmc <- combined1 %>%
  filter(!is.na(campus.x), !is.na(uuid), !is.na(domain), campus.x=="cmc") %>%
  group_by(uuid) %>%
  summarise(numofjournalsused=n_distinct(domain)) %>%
  summarise(avg = mean(numofjournalsused), median=median(numofjournalsused))

usagecmc

## # A tibble: 1 x 2
##       avg median
##   <dbl> <int>
## 1 6.352827     5

# The average CMC student uses 6 different journals, while the median use is 5 journals.

# Pitzer usage
usagepitzer <- combined1 %>%
  filter(!is.na(campus.x), !is.na(uuid), !is.na(domain), campus.x=="pitzer") %>%
  group_by(uuid) %>%
  summarise(numofjournalsused=n_distinct(domain)) %>%
  summarise(avg = mean(numofjournalsused), median=median(numofjournalsused))

usagepitzer

## # A tibble: 1 x 2
##       avg median
##   <dbl> <dbl>
## 1 6.321683     5

```

```
# The average Pitzer student uses 6 different journals, while the median use is 5 journals.
```

```
# Pitzer usage
```

```
usagescripps <- combined1 %>%  
  filter(!is.na(campus.x), !is.na(uuid), !is.na(domain), campus.x=="scrippscollege") %>%  
  group_by(uuid) %>%  
  summarise(numofjournalsused=n_distinct(domain)) %>%  
  summarise(avg = mean(numofjournalsused), median=median(numofjournalsused))
```

```
usagescripps
```

```
## # A tibble: 1 x 2  
##       avg median  
##   <dbl> <int>  
## 1 7.335063     6
```

```
# The average Scripps student uses 7 different journals, while the median use is 6 journals.
```

```
# Mudd Usage
```

```
usagemudd <- combined1 %>%  
  filter(!is.na(campus.x), !is.na(uuid), !is.na(domain), campus.x=="hmc", domain=="jstor") %>%  
  group_by(uuid) %>%  
  summarise(numofjournalsused=n_distinct(domain)) %>%  
  summarise(avg = mean(numofjournalsused), median=median(numofjournalsused))
```

```
usagemudd
```

```
## # A tibble: 1 x 2  
##       avg median  
##   <dbl> <dbl>  
## 1     1     1
```

```
# The average Mudd student uses 5 different journals, while the median use is 5 journals.
```

```
# CGU Usage
```

```
usagecgu <- combined1 %>%  
  filter(!is.na(campus.x), !is.na(uuid), !is.na(domain), campus.x=="cgu") %>%  
  group_by(uuid) %>%  
  summarise(numofjournalsused=n_distinct(domain)) %>%  
  summarise(avg = mean(numofjournalsused), median=median(numofjournalsused))
```

```
usagecgu
```

```
## # A tibble: 1 x 2  
##       avg median  
##   <dbl> <dbl>  
## 1 8.063712     7
```

```
# The average CGU student uses 8 different journals, while the median use is 7 journals.
```

```
usage1 <- combined1 %>%  
  filter(!is.na(campus.x), !is.na(uuid), !is.na(domain), campus.x!="cuc", campus.x!="kecksci", campus.x!="ke  
  group_by(uuid, campus.x) %>%  
  summarise(numofjournalsused=n_distinct(domain))
```

```
ggplot(usage1, aes(x=numofjournalsused, col=campus.x))+geom_freqpoly()+xlab("Number of Different Journals
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

### Distribution of Journal Usage Across the Claremont Colleges

