100 Years of Social Work Research: A Data Science Perspective

Overview of data

The original data were from a search of PsychInfo using Ebsco Host platform (December 23, 2014). The following search operators and limiters were used:

- SO "social work" OR SO "social welfare" OR SO "social casework" OR SO "social services"
- Limiters - Document Type: Journal Article
- Search modes - Boolean/Phrase Interface - EBSCOhost Research Databases
- Search Screen - Advanced Search
- Database - PsycINFO

The search results were exported in the *Generic bibliographic format*. This format is a text (*.txt) file that is processed with the `BibWrangleR` functions.

Initialize workspace and functions for data wrangling

This section processes raw data. This section of code is executed only one time to transform the raw data into an analyzable format. When new data are obtained for this study (i.e., updated search results), this section should be re-run by changing `echo=FALSE` to `echo=TRUE` in the knitr markdown argument.

```
# Clear workspace
rm(list=ls())

# Read BWR functions
source("/Users/beperron/Git/BibWrangleR/functions/piWrangleR.R")
source("/Users/beperron/Git/BibWrangleR/functions/packages.R")

# Set the path where original raw data are stored
setwd("/Users/beperron/Git/SocialWorkResearch")

# Set the working directory to store files created by BWR functions
my.path <- "/Users/beperron/Git/SocialWorkResearch"

# Wrangle the data with the BWR function suite
#piBWR.f(csv=FALSE, path=my.path)
#save(pi.df, file = "piArticles.R")
```

Initialize workspace and functions for analaysis

All the analyses performed involve the data that have been processed with the `BibWrangleR` functions. This section reads the processed data, loads the required packages, and does a quick quality check to ensure that

the same number of articles (i.e., records) contained in the original search match the number of articles in the transformed data.

```
rm(list=ls())
load("piArticles.R")
library(dplyr)
library(ggplot2)
library(gridExtra)

# Dimensions of the data file (Rows X Columns)
dim(pi.df)
```

```
[1] 495415      3
```

```
#Variables of the data file
names(pi.df)
```

```
[1] "attributes" "articleID"  "record"
```

```
# How many unique article titles?  Ebsco Results of most current search is $n=24,314$. Do not proceed w
length(which(pi.df$attributes == "TI"))
```

```
[1] 24314
```

Overall number and name of unique journal titles?

```
unique.titles <- filter(pi.df, attributes == "SO")

#Number of unique titles
length(unique(unique.titles$record))
```

```
## [1] 89
```

```
#Unique titles
unique(unique.titles$record)
```

```
##  [1] "Journal of Ethnic & Cultural Diversity in Social Work: Innovation in Theory, Research & Practi
##  [2] "Journal of Sociology and Social Welfare"
##  [3] "Social Work & Christianity"
##  [4] "Journal of Gerontological Social Work"
##  [5] "Research on Social Work Practice"
##  [6] "Child & Family Social Work"
##  [7] "Australian Social Work"
##  [8] "Social Work with Groups: A Journal of Community and Clinical Practice"
##  [9] "Practice: Social Work in Action"
## [10] "Journal of Gay & Lesbian Social Services: The Quarterly Journal of Community & Clinical Practi
## [11] "Smith College Studies in Social Work"
## [12] "Journal of Social Work Practice"
```

```
## [13] "Social Work in Health Care"
## [14] "Journal of Social Work Education"
## [15] "Children & Schools"
## [16] "Social Work"
## [17] "Child & Adolescent Social Work Journal"
## [18] "Clinical Social Work Journal"
## [19] "International Social Work"
## [20] "Journal of Social Work"
## [21] "Social Work Research"
## [22] "Social Work Education"
## [23] "Journal of Evidence-Based Social Work"
## [24] "Health & Social Work"
## [25] "Affilia: Journal of Women & Social Work"
## [26] "Qualitative Social Work: Research and Practice"
## [27] "Families in Society"
## [28] "Social Work in Mental Health"
## [29] "Ethics and Social Welfare"
## [30] "Journal of Religion & Spirituality in Social Work: Social Thought"
## [31] "Journal of HIV/AIDS & Social Services"
## [32] "Journal of Social Work Practice in the Addictions"
## [33] "British Journal of Social Work"
## [34] "School Social Work Journal"
## [35] "Journal of the Society for Social Work and Research"
## [36] "Journal of Social Work in End-of-Life & Palliative Care"
## [37] "International Journal of Social Welfare"
## [38] "Psychoanalytic Social Work"
## [39] "Administration in Social Work"
## [40] "The Journal of Baccalaureate Social Work"
## [41] "The Scientific Review of Mental Health Practice: Objective Investigations of Controversial and
## [42] "Social Work and Social Sciences Review"
## [43] "Journal of Gay & Lesbian Social Services: Issues in Practice, Policy & Research"
## [44] "Practice"
## [45] "Journal of Educational & Psychological Consultation"
## [46] "Rural Social Work"
## [47] "Journal of Technology in Human Services"
## [48] "Journal of Social Service Research"
## [49] "Journal of Applied Social Sciences"
## [50] "Early Child Development and Care"
## [51] "Computers in Human Services"
## [52] "The Clinical Supervisor"
## [53] "Children and Youth Services Review"
## [54] "Journal of Social Work Research and Evaluation"
## [55] "General Hospital Psychiatry"
## [56] "Canadian Journal on Aging"
## [57] "Social Casework"
## [58] "Journal of Multicultural Social Work"
## [59] "Journal of Analytic Social Work"
## [60] "Maatskaplike Werk/Social Work"
## [61] "Issues in Social Work Education"
## [62] "Journal of Teaching in Social Work"
## [63] "Social Work Research & Abstracts"
## [64] "Journal of Social Work & Human Sexuality"
## [65] "Journal of Independent Social Work"
## [66] "Employee Assistance Quarterly"
```

```
## [67] "Behavior Modification"
## [68] "Indian Journal of Social Work"
## [69] "Indian Journal of Psychiatric Social Work"
## [70] "British Journal of Psychiatric Social Work"
## [71] "Social Work in Education"
## [72] "Pediatric Social Work"
## [73] "Journal of Social Welfare"
## [74] "School Social Work Quarterly"
## [75] "Social Work Today"
## [76] "Journal of Psychiatric Social Work"
## [77] "Medical Social Work"
## [78] "Jewish Social Services Quarterly"
## [79] "Proceedings of the National Conference of Social Work"
## [80] "Journal of Social Casework"
## [81] "Social Work Yearbook"
## [82] "Social Work Technique"
## [83] "Journal of Social Work Process"
## [84] "Pennsylvania Social Work"
## [85] "International Conference of Social Work"
## [86] "Eugenics & Social Welfare Bull."
## [87] "New York State Department of Social Welfare, Division Publication"
## [88] "University of Washington Publications: Social Services"
## [89] "Eugenics and Social Welfare Bulletin"
```

Number of unique journal titles by year

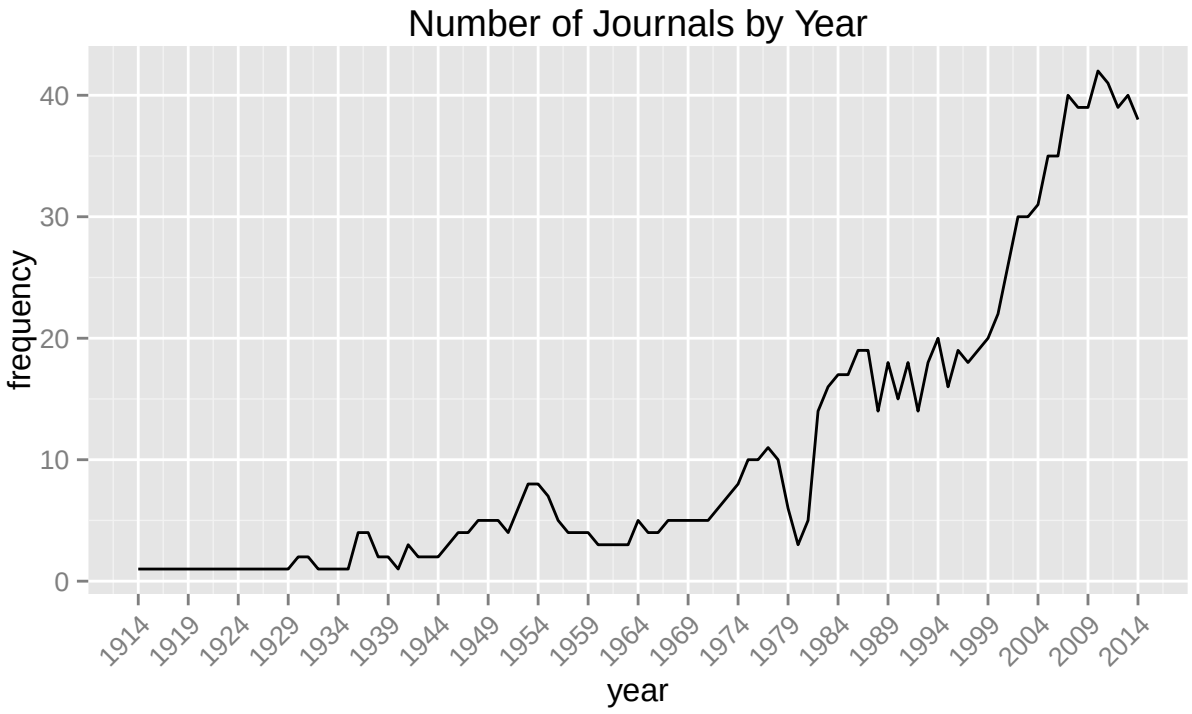
```
journals.year <- tbl_df(pi.df)
year <- journals.year %>%
        filter(attributes == "YR") %>%
        select(id = articleID, year = record)

journals <- journals.year %>%
        filter(attributes == "SO") %>%
        select(id = articleID, journal.title = record)

n.journals.year <- journals %>%
        left_join(year) %>%
        group_by(year) %>%
        distinct(journal.title) %>%
        summarise(n = n())


journal.count <- ggplot(n.journals.year, aes(as.numeric(year), y=n, group=1)) +
    geom_line(colour="black") +
    #geom_point(colour="red") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    xlab("year") +
    ylab("frequency") +
    ggtitle("Number of Journals by Year") +
    scale_x_continuous(breaks=seq(1914, 2014, 5))

journal.count
```



What journals published the most number of articles

```
n.so.yr <- filter(pi.df, attributes == "SO" | attributes == "YR")
n.so <- filter(pi.df, attributes == "SO") %>% mutate(title = record) %>% select(-attributes, -record)
n.yr <- filter(pi.df, attributes == "YR") %>% mutate(year = record ) %>% select(-attributes, -record)
n.so.yr <- left_join(n.so, n.yr) %>%
    group_by(title) %>%
    summarise(
        first = min(year),
        last = max(year),
        n.to.date = n()
        ) %>%
    arrange(desc(n.to.date))
```

```
## Joining by: "articleID"
```

```
#10 highest number of publications
head(n.so.yr, 10)
```

```
## Source: local data frame [10 x 4]
##
##                                 title first last n.to.date
## 1                         Social Work  1948 2014      1866
## 2         British Journal of Social Work  1971 2014      1456
## 3                   Families in Society  1990 2014      1211
```

```
## 4  Journal of Gerontological Social Work  1981 2014      1188
## 5             Social Work in Health Care  1975 2014      1171
## 6                         Social Casework  1950 1989      1095
## 7   Smith College Studies in Social Work  1930 2014      1075
## 8           Clinical Social Work Journal  1973 2014      1068
## 9       Research on Social Work Practice  1991 2014       986
## 10                  Health & Social Work  1976 2014       901
```

What is the lifespan of journals?

```
#10 longest running journals
longest.running <- n.so.yr %>%
       mutate(last = as.numeric(last), first = as.numeric(first),year.diff = last - first) %>%
       arrange(desc(year.diff)) %>%
       select(title, first, last, year.diff)
head(longest.running, 10)
```

```
## Source: local data frame [10 x 4]
##
##                                    title first last year.diff
## 1     Smith College Studies in Social Work  1930 2014        84
## 2                              Social Work  1948 2014        66
## 3                   Journal of Social Work  1964 2014        50
## 4            Indian Journal of Social Work  1941 1986        45
## 5           British Journal of Social Work  1971 2014        43
## 6             Clinical Social Work Journal  1973 2014        41
## 7  Journal of Sociology and Social Welfare  1974 2014        40
## 8               Social Work in Health Care  1975 2014        39
## 9                         Social Casework  1950 1989        39
## 10                  Health & Social Work  1976 2014        38
```

```
#10 shortest running journals
shortest.running <- longest.running %>% arrange(year.diff, first, last)
head(shortest.running, 20)
```

```
## Source: local data frame [20 x 4]
##
##                                                                title first
## 1                              Eugenics and Social Welfare Bulletin  1918
## 2                            International Conference of Social Work  1928
## 3             University of Washington Publications: Social Services  1929
## 4                                        Pennsylvania Social Work  1936
## 5                                   Journal of Social Work Process  1937
## 6  New York State Department of Social Welfare, Division Publication  1937
## 7                                            Social Work Yearbook  1947
## 8                                               Social Work Today  1978
## 9                                  School Social Work Quarterly  1979
## 10                                        Pediatric Social Work  1984
## 11                                  Canadian Journal on Aging  1987
## 12                                      Behavior Modification  1989
```

```
## 13                                    Employee Assistance Quarterly  1989
## 14                                 Children and Youth Services Review  1991
## 15                                        General Hospital Psychiatry  1991
## 16                                        Computers in Human Services  1994
## 17                                            The Clinical Supervisor  1994
## 18                                 Journal of Applied Social Sciences  1994
## 19                                 Journal of Teaching in Social Work  1994
## 20                                   Early Child Development and Care  1995
## Variables not shown: last (dbl), year.diff (dbl)
```

Number of articles per year

```
n.articles.year <- filter(pi.df, attributes == "YR")
year.split <- split(n.articles.year, n.articles.year$record)
year.count <- unlist(lapply(year.split, nrow))
year.count <- year.count[order(names(year.count))]
years <- names(year.count)

df <- data.frame(years, year.count)
rownames(df) <- NULL

head(df, 10)
```

```
##    years year.count
## 1   1914          1
## 2   1915          1
## 3   1918          1
## 4   1928          2
## 5   1929          1
## 6   1930         16
## 7   1931         15
## 8   1932         13
## 9   1933         15
## 10  1934          7
```

```
tail(df,10)
```

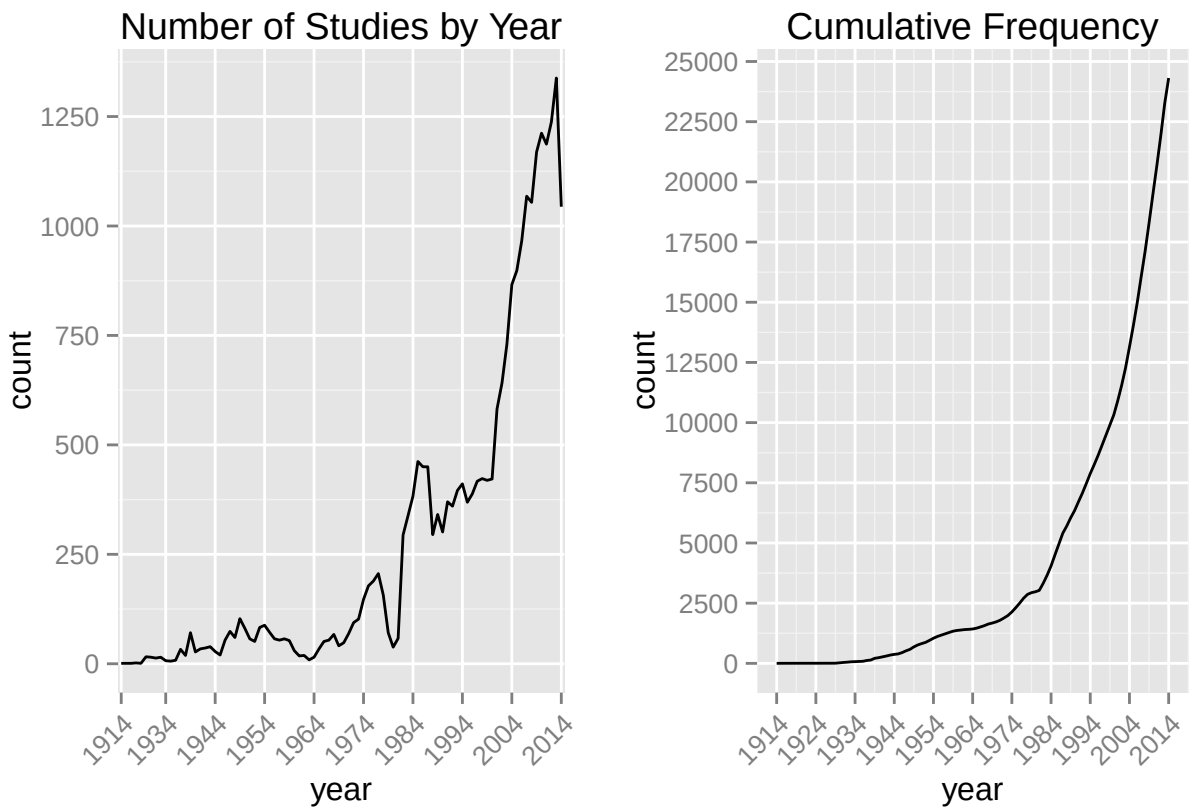
```
##    years year.count
## 81  2005        898
## 82  2006        966
## 83  2007       1068
## 84  2008       1054
## 85  2009       1169
## 86  2010       1212
## 87  2011       1187
## 88  2012       1238
## 89  2013       1338
## 90  2014       1044
```

```
plot.article.count <- ggplot(df, aes(as.factor(years), y= year.count, group=1)) +
    geom_line(colour="black") +
    #geom_point(colour="red") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    xlab("year") +
    ylab("count") +
    ggtitle("Number of Studies by Year") +
    scale_x_discrete(breaks=c(seq(1914, 2014, 10))) +
    scale_y_continuous(breaks = c(seq(0, 2000, 250)))

df$years <- as.numeric(as.character(df$years))
plot.article.cumulative <- ggplot(df, aes(x = years, y = cumsum(year.count))) +
    geom_line() +
    theme(axis.text.x = element_text(angle=45, hjust=1)) +
    scale_x_continuous(breaks=pretty(df$years)) +
    xlab("year") +
    ylab("count") +
    scale_x_continuous(breaks = c(seq(1914,2014,10))) +
    scale_y_continuous(breaks = c(seq(0, 25000, 2500))) +
    ggtitle("Cumulative Frequency")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale
```

```
grid.arrange(plot.article.count, plot.article.cumulative, ncol=2)
```



Topic areas (by Subject Terms)

It is easy to explore some of the different fields within the PsychInfo data frame. For example, each record has one or more subject terms (from the article keywords). The total number, unique number, and most frequently occuring key words can be easily computed.

```
df.2 <- filter(pi.df, attributes == "SU")


subject.terms <- stringr::str_split(df.2$record, pattern = ";")
subject.terms <- unlist(lapply(subject.terms, function(x) gsub(" ", "", x)))
subject.terms.total <- length(unlist(lapply(subject.terms, function(x) gsub(" ", "", x))))
subject.terms.unique <- length(unique(subject.terms))

subject.terms.l <- list(subject.terms.total = subject.terms.total,
                        subject.terms.unique = subject.terms.unique)

most.frequent <- as.data.frame(table(subject.terms))



most.frequent <- arrange(most.frequent, desc(Freq))
most.frequent.t <- head(most.frequent, 50)

print(subject.terms.l)
```

```
$subject.terms.total
[1] 92161

$subject.terms.unique
[1] 3037
```

```
print(most.frequent.t)
```

```
                subject.terms Freq
1               SocialCasework 5794
2                SocialWorkers 2933
3          SocialWorkEducation 1696
4               SocialServices 1139
5                 ChildWelfare  811
6                SocialSupport  602
7             CommunityServices  572
8                       Family  572
9                   Caregivers  571
10                  ChildAbuse  571
11              MentalDisorders  500
12                 HumanFemales  493
13                    DrugAbuse  488
14              FamilyRelations  478
15                        Aging  474
16                          HIV  471
17                  FosterCare  470
```

```
18                      Blacks  445
19         MentalHealthServices  441
20           HealthCareServices  440
21                 MentalHealth  438
22               CopingBehavior  403
23                 Intervention  400
24           GroupPsychotherapy  393
25  PsychotherapeuticProcesses  390
26                    Treatment  359
27              GroupCounseling  357
28             DomesticViolence  345
29           ProfessionalEthics  343
30               DecisionMaking  335
31           HealthCareDelivery  333
32        EvidenceBasedPractice  327
33                         AIDS  325
34                FamilyTherapy  324
35              Experimentation  315
36            MaleHomosexuality  315
37                      Parents  315
38                Psychotherapy  315
39       GovernmentPolicyMaking  314
40                  Immigration  311
41                 Spirituality  303
42           DrugRehabilitation  302
43 WelfareServices(Government)  297
44                      Mothers  287
45           ProtectiveServices  287
46     HealthPersonnelAttitudes  284
47                  SexualAbuse  280
48            AtRiskPopulations  279
49              MajorDepression  276
50                   Lesbianism  272
```

Most Frequent Subject Terms Over Time (by 5 years)

```
decade <- filter(pi.df, attributes == "YR") %>%
    mutate(year = as.numeric(record)) %>% select(-record, -attributes)

decade$year <- cut(decade$year, breaks = 20, labels = c(1:20))

keywords <- pi.df %>%
        filter(attributes == "SU") %>%
        select(articleID = articleID, keywords = record)

keywords.decade <- keywords %>%
         left_join(decade)
```

```
## Joining by: "articleID"
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
keywords.data.split <- dlply(keywords.decade, .(year))
detach(package:plyr)

terms.f <- function(x){
    split.terms <- stringr::str_split(x[,"keywords"], pattern =";")
    clean.terms <- lapply(split.terms, function(x) gsub(" ", "", x))
    }

keywords.decade <- lapply(keywords.data.split, terms.f)
keywords.decade <- lapply(keywords.decade, unlist)
lapply(keywords.decade, function(x) length(unique(x)))
```

```
## $`5`
## [1] 2
##
## $`6`
## [1] 2
##
## $`7`
## [1] 2
##
## $`8`
## [1] 3
##
## $`9`
## [1] 4
##
## $`10`
## [1] 44
##
## $`11`
## [1] 337
##
## $`12`
## [1] 512
##
## $`13`
```

```
## [1] 796
##
## $`14`
## [1] 902
##
## $`15`
## [1] 1179
##
## $`16`
## [1] 1190
##
## $`17`
## [1] 1384
##
## $`18`
## [1] 1708
##
## $`19`
## [1] 1941
##
## $`20`
## [1] 2003
```

```
temp <- lapply(keywords.decade, function(x) data.frame(table(x)))
temp <- lapply(temp, function(x) arrange(x, desc(Freq)))
lapply(temp, function(x) head(x,10))
```

```
## $`5`
##               x Freq
## 1 ChildGuidance    2
## 2        Agency    1
##
## $`6`
##               x Freq
## 1        Agency    1
## 2 ChildGuidance    1
##
## $`7`
##               x Freq
## 1        Agency   11
## 2 ChildGuidance    2
##
## $`8`
##                 x Freq
## 1          Agency    9
## 2   ChildGuidance    5
## 3 SmallBusinesses    1
##
## $`9`
##                       x Freq
## 1                Agency    9
## 2         ChildGuidance    2
## 3 EmotionalDisturbances    1
## 4       SocialGroupWork    1
```

```
##
## $`10`
##                         x Freq
## 1                 Clients    3
## 2  EmotionalDisturbances    3
## 3        FamilyRelations    2
## 4          FamilyTherapy    2
## 5             Infidelity    2
## 6          SocialWorkers    2
## 7     AntisocialBehavior    1
## 8             ChildAbuse    1
## 9          ChildGuidance    1
## 10       ChildPsychiatry    1
##
## $`11`
##                            x Freq
## 1            SocialCasework   99
## 2                 Treatment   25
## 3         CommunityServices   20
## 4             FamilyTherapy   15
## 5                    Family   14
## 6  InterpersonalInteraction   12
## 7           FamilyRelations   11
## 8       PsychiatricPatients   11
## 9        ChildPsychotherapy   10
## 10                  Clinics   10
##
## $`12`
##                              x Freq
## 1              SocialCasework  145
## 2               SocialWorkers   39
## 3             FamilyRelations   30
## 4           CommunityServices   25
## 5               FamilyTherapy   25
## 6         PsychiatricPatients   19
## 7          GroupPsychotherapy   17
## 8        ParentChildRelations   16
## 9  PsychotherapeuticProcesses   15
## 10        SocioeconomicStatus   14
##
## $`13`
##                               x Freq
## 1               SocialCasework  212
## 2                SocialWorkers   99
## 3              FamilyRelations   37
## 4                FamilyTherapy   35
## 5                 HumanFemales   34
## 6   PsychotherapeuticProcesses   34
## 7          SocialWorkEducation   34
## 8                      Parents   30
## 9  PsychotherapeuticTechniques   30
## 10                  Counseling   28
##
## $`14`
```

```
##                   x Freq
## 1    SocialCasework  315
## 2     SocialWorkers  129
## 3   GroupCounseling   76
## 4    SocialServices   48
## 5      HumanFemales   45
## 6     SocialSupport   44
## 7  LiteratureReview   43
## 8        ChildAbuse   37
## 9     FamilyTherapy   37
## 10  FamilyRelations   34
##
## $`15`
##                     x Freq
## 1      SocialCasework  566
## 2       SocialWorkers  203
## 3      SocialServices  111
## 4     GroupCounseling  105
## 5       SocialSupport   87
## 6  GroupPsychotherapy   81
## 7       FamilyTherapy   74
## 8          ChildAbuse   68
## 9      CopingBehavior   62
## 10       HumanFemales   59
##
## $`16`
##                  x Freq
## 1   SocialCasework  411
## 2    SocialWorkers  174
## 3   SocialServices   83
## 4  FamilyRelations   78
## 5       Caregivers   75
## 6  MentalDisorders   74
## 7       ChildAbuse   72
## 8    SocialSupport   61
## 9             AIDS   60
## 10 GroupCounseling   55
##
## $`17`
##                      x Freq
## 1       SocialCasework  419
## 2        SocialWorkers  151
## 3       SocialServices   91
## 4           ChildAbuse   89
## 5  SocialWorkEducation   81
## 6        SocialSupport   70
## 7      FamilyRelations   65
## 8               Blacks   64
## 9                  HIV   62
## 10        ChildWelfare   58
##
## $`18`
##                      x Freq
## 1       SocialCasework  699
```

```
## 2        SocialWorkers  399
## 3  SocialWorkEducation  198
## 4       SocialServices  186
## 5         ChildWelfare  140
## 6            DrugAbuse  107
## 7    MaleHomosexuality  103
## 8           Caregivers  102
## 9        SocialSupport   93
## 10          FosterCare   88
##
## $`19`
##                      x Freq
## 1       SocialCasework 1377
## 2        SocialWorkers  824
## 3  SocialWorkEducation  566
## 4         ChildWelfare  291
## 5       SocialServices  281
## 6                Aging  180
## 7                  HIV  169
## 8               Family  164
## 9           Caregivers  163
## 10        Intervention  162
##
## $`20`
##                      x Freq
## 1       SocialCasework 1550
## 2        SocialWorkers  907
## 3  SocialWorkEducation  676
## 4       SocialServices  339
## 5         ChildWelfare  282
## 6         Intervention  201
## 7                Aging  189
## 8         MentalHealth  183
## 9               Family  179
## 10          Caregivers  154
```

Author defined keywords

It is easy to explore some of the different fields within the PsychInfo data frame. For example, each record has one or more subject terms (from the article keywords). The total number, unique number, and most frequently occuring key words can be easily computed.

```
kp.df <- filter(pi.df, attributes == "KP")


subject.terms <- stringr::str_split(kp.df$record, pattern = ";")
subject.terms <- unlist(lapply(subject.terms, function(x) gsub(" ", "", x)))
subject.terms.total <- length(unlist(lapply(subject.terms, function(x) gsub(" ", "", x))))
subject.terms.unique <- length(unique(subject.terms))

subject.terms.l <- list(subject.terms.total = subject.terms.total,
                        subject.terms.unique = subject.terms.unique)
```

```r
most.frequent <- as.data.frame(table(subject.terms))



most.frequent <- arrange(most.frequent, desc(Freq))
most.frequent.t <- head(most.frequent, 50)

print(subject.terms.l)
```

```
$subject.terms.total
[1] 102493

$subject.terms.unique
[1] 46899
```

```r
print(most.frequent.t)
```

```
               subject.terms Freq
1               socialworkers 1766
2                  socialwork 1757
3         socialworkeducation  756
4          socialworkpractice  538
5              socialservices  340
6          socialworkstudents  314
7                mentalhealth  304
8                    children  283
9                childwelfare  255
10   CHILDHOODANDADOLESCENCE  236
11                       HIV  231
12             socialsupport  226
13               riskfactors  200
14               spirituality  199
15            decisionmaking  188
16                fostercare  187
17                     aging  179
18          domesticviolence  176
19             socialjustice  176
20                TECHNIQUES  175
21              intervention  174
22               adolescents  173
23               METHODOLOGY  173
24                SOCIALWORK  173
25             CHILDGUIDANCE  167
26    conferencepresentation  165
27                     India  163
28                depression  161
29           childprotection  157
30                   poverty  154
31                      AIDS  152
32                     CHILD  149
33     evidencebasedpractice  149
34             substanceabuse  149
```

```
35               olderadults  148
36                  families  147
37                    family  147
38                caregivers  145
39      mentalhealthservices  144
40          literaturereview  143
41                    Israel  142
42            SOCIALCASEWORK  140
43                 wellbeing  138
44                  GUIDANCE  136
45                   parents  135
46 implicationsforsocialwork  133
47              testvalidity  130
48                     women  125
49             psychometrics  124
50             psychotherapy  124
```

Most Frequent Author Keywords

```
decade <- filter(pi.df, attributes == "YR") %>%
    mutate(year = as.numeric(record)) %>% select(-record, -attributes)

decade$year <- cut(decade$year, breaks = 20, labels = c(1:20))

keywords <- pi.df %>%
        filter(attributes == "KP") %>%
        select(articleID = articleID, keywords = record)

keywords.decade <- keywords %>%
         left_join(decade)
```

```
## Joining by: "articleID"
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
keywords.data.split <- dlply(keywords.decade, .(year))
detach(package:plyr)
```

```
terms.f <- function(x){
    split.terms <- stringr::str_split(x[,"keywords"], pattern =";")
    clean.terms <- lapply(split.terms, function(x) gsub(" ", "", x))
    }

keywords.decade <- lapply(keywords.data.split, terms.f)
keywords.decade <- lapply(keywords.decade, unlist)
lapply(keywords.decade, function(x) length(unique(x)))
```

```
## $`3`
## [1] 8
##
## $`4`
## [1] 171
##
## $`5`
## [1] 285
##
## $`6`
## [1] 278
##
## $`7`
## [1] 714
##
## $`8`
## [1] 711
##
## $`9`
## [1] 835
##
## $`10`
## [1] 383
##
## $`11`
## [1] 573
##
## $`12`
## [1] 988
##
## $`13`
## [1] 1705
##
## $`14`
## [1] 2222
##
## $`15`
## [1] 4033
##
## $`16`
## [1] 3702
##
## $`17`
## [1] 4068
```

```
##
## $`18`
## [1] 9760
##
## $`19`
## [1] 13125
##
## $`20`
## [1] 13587
```

```
temp <- lapply(keywords.decade, function(x) data.frame(table(x)))
temp <- lapply(temp, function(x) arrange(x, desc(Freq)))
lapply(temp, function(x) head(x,10))
```

```
## $`3`
##                                x Freq
## 1 SOCIALFUNCTIONSOFTHEINDIVIDUAL    2
## 2                      COMMUNITY    1
## 3                 COMMUNITYSPIRIT    1
## 4                        COUNTRY    1
## 5                    LIMITATIONS    1
## 6                SOCIALADJUSTMENT    1
## 7                         SPIRIT    1
## 8                           WORK    1
##
## $`4`
##                                 x Freq
## 1         CHILDHOODANDADOLESCENCE   38
## 2  SOCIALFUNCTIONSOFTHEINDIVIDUAL   25
## 3                           CHILD   19
## 4       NERVOUSANDMENTALDISORDERS   12
## 5                     DELINQUENCY    8
## 6                          FAMILY    8
## 7                  CHILDABILITIES    6
## 8  MOTHERATTITUDEANDBREASTFEEDING    6
## 9                     PERSONALITY    5
## 10                     ADJUSTMENT    4
##
## $`5`
##                                       x Freq
## 1                 CHILDHOODANDADOLESCENCE   56
## 2  GENERALSOCIALPROCESSES(INCL.ESTHETICS)   53
## 3                                   CHILD   33
## 4                     FUNCTIONALDISORDERS   30
## 5                           SOCIALABILITY   29
## 6                                    WORK   27
## 7                                GUIDANCE   20
## 8                                CASEWORK   16
## 9          CHILD(MALADJUSTMENTANDTHERAPY)   16
## 10                               ATTITUDE   15
##
## $`6`
##                                       x Freq
## 1                 CHILDHOODANDADOLESCENCE   99
```

```
## 2                 CHILD(IV.MALADJUSTMENT   51
## 3                                THERAPY)   51
## 4                                GUIDANCE   43
## 5                    FUNCTIONALDISORDERS   40
## 6                                  CHILD   30
## 7          CHILD(MALADJUSTMENTANDTHERAPY)   28
## 8                                 SOCIAL   24
## 9                              ADJUSTMENT   22
## 10 GENERALSOCIALPROCESSES(INCL.ESTHETICS)   21
##
## $`7`
##                          x Freq
## 1           SOCIALCASEWORK   78
## 2            CHILDGUIDANCE   63
## 3  CHILDHOODANDADOLESCENCE   43
## 4                   SOCIAL   30
## 5         TREATMENTMETHODS   30
## 6                   CHILD   29
## 7                GUIDANCE   23
## 8      FUNCTIONALDISORDERS   22
## 9              COUNSELING   21
## 10           PSYCHOTHERAPY   21
##
## $`8`
##                  x Freq
## 1        SOCIALWORK  111
## 2       METHODOLOGY   76
## 3        TECHNIQUES   76
## 4              CASE   61
## 5     CHILDGUIDANCE   53
## 6  TREATMENTMETHODS   50
## 7        COUNSELING   41
## 8    SOCIALCASEWORK   35
## 9          GUIDANCE   32
## 10           FAMILY   30
##
## $`9`
##                     x Freq
## 1          TECHNIQUES   78
## 2         METHODOLOGY   77
## 3       SOCIALWELFARE   58
## 4    TREATMENTMETHODS   42
## 5          SOCIALWORK   39
## 6       CHILDGUIDANCE   37
## 7              FAMILY   34
## 8          COUNSELING   31
## 9   CRIME&DELINQUENCY   25
## 10 SOCIALINSTITUTIONS   25
##
## $`10`
##                          x Freq
## 1     CASESTUDIES&CASEWORK   14
## 2           SOCIALWELFARE   14
## 3           PSYCHOTHERAPY   11
```

```
## 4         BEHAVIORPROBLEMS   10
## 5                CHILDHOOD    6
## 6  MARRIAGE&FAMILYPROBLEMS    6
## 7           SOCIALCASEWORK    6
## 8              ALCOHOLISM    5
## 9        CRIME&DELINQUENCY    5
## 10                 FAMILY    5
##
## $`11`
##                      x Freq
## 1              clients    7
## 2       socialcasework    6
## 3                India    5
## 4       SOCIALCASEWORK    5
## 5           socialwork    5
## 6  COUNSELING&GUIDANCE    4
## 7        socialworkers    4
## 8  CHILDGUIDANCECLINIC    3
## 9                  ego    3
## 10               INDIA    3
##
## $`12`
##                            x Freq
## 1              socialworkers   19
## 2                      India   18
## 3                  socialwork    8
## 4                  casereport    4
## 5                 drugaddicts    3
## 6  graduatesocialworkstudents    3
## 7                grouptherapy    3
## 8                      Israel    3
## 9            literaturereview    3
## 10                     racism    3
##
## $`13`
##                              x Freq
## 1                        India   24
## 2                socialworkers   22
## 3             literaturereview    9
## 4                          aged    8
## 5            socialworkstudents    8
## 6                     children    5
## 7     implicationsforsocialwork    5
## 8                socialcasework    5
## 9  implicationsforsocialworkers    4
## 10    malevsfemalesocialworkers    4
##
## $`14`
##                            x Freq
## 1              socialworkers   63
## 2                      India   59
## 3           literaturereview   41
## 4  implicationsforsocialwork   19
## 5                  children   17
```

```
## 6                    elderly   15
## 7                      aged   13
## 8                 socialwork   11
## 9          socialworkpractice    9
## 10               adolescents    7
##
## $`15`
##                               x Freq
## 1                 socialworkers  113
## 2         conferencepresentation   57
## 3     implicationsforsocialwork   44
## 4                         India   35
## 5                    casereport   32
## 6               literaturereview   28
## 7                        elderly   27
## 8         socialworkimplications   25
## 9  implicationsforsocialworkers   24
## 10                  adolescents   22
##
## $`16`
##                             x Freq
## 1               socialworkers   65
## 2     conferencepresentation   47
## 3           literaturereview   34
## 4  implicationsforsocialwork   31
## 5                     Israel   31
## 6                    England   27
## 7                     Canada   17
## 8                    elderly   17
## 9                    clients   16
## 10    socialworkimplications   16
##
## $`17`
##                                   x Freq
## 1                     socialworkers   73
## 2             conferencepresentation   55
## 3                             Israel   38
## 4  implicationsforsocialworkpractice   25
## 5                   literaturereview   25
## 6          implicationsforsocialwork   24
## 7                            England   23
## 8                         casereport   22
## 9       implicationsforsocialworkers   15
## 10                socialworkstudents   14
##
## $`18`
##                       x Freq
## 1           socialwork  312
## 2         socialworkers  282
## 3   socialworkpractice  116
## 4  socialworkeducation   87
## 5              children   66
## 6          mentalhealth   59
## 7       socialservices   53
```

```
## 8               gaymen   48
## 9        socialsupport   47
## 10          depression   46
##
## $`19`
##                      x Freq
## 1           socialwork  611
## 2        socialworkers  486
## 3  socialworkeducation  276
## 4   socialworkpractice  195
## 5   socialworkstudents  109
## 6       socialservices  107
## 7                  HIV  100
## 8         childwelfare   93
## 9             children   87
## 10        mentalhealth   80
##
## $`20`
##                      x Freq
## 1           socialwork  800
## 2        socialworkers  638
## 3  socialworkeducation  385
## 4   socialworkpractice  206
## 5       socialservices  176
## 6         mentalhealth  156
## 7   socialworkstudents  123
## 8         childwelfare  121
## 9                  HIV  110
## 10         riskfactors  110
```

Location

It is easy to explore some of the different fields within the PsychInfo data frame. For example, each record has one or more subject terms (from the article keywords). The total number, unique number, and most frequently occuring key words can be easily computed.

```
LO.df <- filter(pi.df, attributes == "LO")


subject.terms <- stringr::str_split(LO.df$record, pattern = ";")
subject.terms <- unlist(lapply(subject.terms, function(x) gsub(" ", "", x)))
subject.terms.total <- length(unlist(lapply(subject.terms, function(x) gsub(" ", "", x))))
subject.terms.unique <- length(unique(subject.terms))

subject.terms.l <- list(subject.terms.total = subject.terms.total,
                        subject.terms.unique = subject.terms.unique)

most.frequent <- as.data.frame(table(subject.terms))



most.frequent <- arrange(most.frequent, desc(Freq))
```

```r
most.frequent.t <- head(most.frequent, 50)

print(subject.terms.l)
```

```
$subject.terms.total
[1] 11076

$subject.terms.unique
[1] 204
```

```r
print(most.frequent.t)
```

```
     subject.terms Freq
1               US 5308
2    UnitedKingdom  696
3        Australia  558
4           Canada  542
5           Israel  448
6          England  388
7            India  232
8           Sweden  227
9         HongKong  216
10           China  158
11     SouthAfrica  115
12          Norway  101
13      NewZealand   97
14        Scotland   85
15           Wales   84
16         Ireland   83
17         Germany   72
18         Finland   70
19     Netherlands   61
20         Denmark   52
21 NorthernIreland   52
22           Spain   47
23    GreatBritain   45
24           Italy   44
25           Japan   42
26         Belgium   41
27       Singapore   41
28          Taiwan   41
29          Africa   38
30          Mexico   34
31          Europe   31
32          France   31
33          Greece   27
34          Russia   27
35           Ghana   24
36           Korea   24
37        Portugal   24
38         Romania   23
39          Brazil   22
```

```
40        Thailand   22
41          Uganda   22
42         Hungary   21
43            Asia   18
44         Nigeria   18
45      PuertoRico   18
46      SouthKorea   18
47         Vietnam   18
48         Austria   17
49     Switzerland   17
50           Kenya   16
```

```
most.frequent
```

```
                     subject.terms Freq
1                               US 5308
2                    UnitedKingdom  696
3                        Australia  558
4                           Canada  542
5                           Israel  448
6                          England  388
7                            India  232
8                           Sweden  227
9                         HongKong  216
10                           China  158
11                      SouthAfrica  115
12                           Norway  101
13                       NewZealand   97
14                         Scotland   85
15                            Wales   84
16                          Ireland   83
17                          Germany   72
18                          Finland   70
19                      Netherlands   61
20                          Denmark   52
21                 NorthernIreland   52
22                            Spain   47
23                     GreatBritain   45
24                            Italy   44
25                            Japan   42
26                          Belgium   41
27                        Singapore   41
28                           Taiwan   41
29                           Africa   38
30                           Mexico   34
31                           Europe   31
32                           France   31
33                           Greece   27
34                           Russia   27
35                            Ghana   24
36                            Korea   24
37                         Portugal   24
38                          Romania   23
39                           Brazil   22
```

40	Thailand	22
41	Uganda	22
42	Hungary	21
43	Asia	18
44	Nigeria	18
45	PuertoRico	18
46	SouthKorea	18
47	Vietnam	18
48	Austria	17
49	Switzerland	17
50	Kenya	16
51	Croatia	15
52	Chile	14
53	Ethiopia	14
54	NorthAmerica	14
55	Poland	14
56	Botswana	13
57	Iran	13
58	Turkey	13
59	Zambia	13
60	Malaysia	12
61	Slovenia	11
62	CzechRepublic	10
63	Zimbabwe	10
64	Bangladesh	9
65	Iceland	9
66	Tanzania	9
67	Guatemala	8
68	Lithuania	8
69	Luxembourg	8
70	Nepal	8
71	Peru	8
72	Philippines	8
73	Argentina	7
74	Bulgaria	7
75	Caribbean	7
76	Colombia	7
77	Cyprus	7
78	Egypt	7
79	ElSalvador	7
80	Georgia	7
81	Jordan	7
82	Pakistan	7
83	Palestine	7
84	Rwanda	7
85	USSR	7
86	Albania	6
87	Estonia	6
88	SaudiArabia	6
89	SriLanka	6
90	TrinidadandTobago	6
91	Ukraine	6
92	UnitedArabEmirates	6
93	Afghanistan	5

94	Cambodia	5
95	Cuba	5
96	DominicanRepublic	5
97	Ecuador	5
98	Indonesia	5
99	Latvia	5
100	Lebanon	5
101	Mongolia	5
102	Slovakia	5
103	Somalia	5
104	Barbados	4
105	Cameroon	4
106	DemocraticRepublicofCongo	4
107	Kuwait	4
108	Lesotho	4
109	Moldova	4
110	Nicaragua	4
111	Oceania/PacificIslands	4
112	SierraLeone	4
113	SouthAmerica	4
114	Azerbaijan	3
115	Bolivia	3
116	Bosnia-Herzegovina	3
117	CentralAmerica	3
118	CostaRica	3
119	Czechoslovakia	3
120	Guyana	3
121	Haiti	3
122	Honduras	3
123	Iraq	3
124	Jamaica	3
125	Kazakhstan	3
126	Liberia	3
127	Malawi	3
128	Malta	3
129	MarshallIslands	3
130	Mauritius	3
131	Tajikistan	3
132	Yugoslavia	3
133	Appalachia	2
134	Bermuda	2
135	Bhutan	2
136	EasternEurope	2
137	Fiji	2
138	Gambia	2
139	Kyrgyzstan	2
140	MiddleEast	2
141	Mozambique	2
142	Myanmar	2
143	NewCaledonia	2
144	Palau	2
145	Panama	2
146	PapuaNewGuinea	2
147	Paraguay	2

148	RepublicofSerbia	2
149	Samoa	2
150	Scandinavia	2
151	Swaziland	2
152	Tonga	2
153	Uruguay	2
154	WesternEurope	2
155	Angola	1
156	Armenia	1
157	Bahamas	1
158	Bahrain	1
159	BalticStates	1
160	Belarus	1
161	Brunei	1
162	Burundi	1
163	ChannelIslands	1
164	CommonwealthofIndependentStates	1
165	Comoros	1
166	CookIslands	1
167	Eritrea	1
168	FrenchPolynesia	1
169	Gabon	1
170	Grenada	1
171	Guinea	1
172	IvoryCoast	1
173	Kiribati	1
174	Laos	1
175	LatinAmerica	1
176	Liechtenstein	1
177	Macau	1
178	Macedonia	1
179	Madagascar	1
180	Maldives	1
181	Micronesia(FederatedStatesof)	1
182	Morocco	1
183	Namibia	1
184	Nauru	1
185	Niue	1
186	NorthKorea	1
187	Oman	1
188	Qatar	1
189	RepublicofCongo	1
190	Senegal	1
191	SerbiaandMontenegro	1
192	SlovakRepublic	1
193	SolomonIslands	1
194	StKitts	1
195	Sudan	1
196	Togo	1
197	Tuvalu	1
198	USVirginIslands	1
199	Uzbekistan	1
200	Vanuatu	1
201	Venezuela	1

```
202                       WestBank    1
203                     WestIndies    1
204                          Yemen    1
```

Location over time

```
decade <- filter(pi.df, attributes == "YR") %>%
    mutate(year = as.numeric(record)) %>% select(-record, -attributes)

decade$year <- cut(decade$year, breaks = 20, labels = c(1:20))

keywords <- pi.df %>%
        filter(attributes == "LO") %>%
        select(articleID = articleID, keywords = record)

keywords.decade <- keywords %>%
         left_join(decade)
```

```
## Joining by: "articleID"
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
keywords.data.split <- dlply(keywords.decade, .(year))
detach(package:plyr)

terms.f <- function(x){
    split.terms <- stringr::str_split(x[,"keywords"], pattern =";")
    clean.terms <- lapply(split.terms, function(x) gsub(" ", "", x))
    }

keywords.decade <- lapply(keywords.data.split, terms.f)
keywords.decade <- lapply(keywords.decade, unlist)
lapply(keywords.decade, function(x) length(unique(x)))
```

```
## $`10`
## [1] 1
##
```

```
## $`11`
## [1] 4
##
## $`12`
## [1] 8
##
## $`13`
## [1] 12
##
## $`14`
## [1] 19
##
## $`15`
## [1] 28
##
## $`16`
## [1] 37
##
## $`17`
## [1] 44
##
## $`18`
## [1] 94
##
## $`19`
## [1] 118
##
## $`20`
## [1] 179
```

```
temp <- lapply(keywords.decade, function(x) data.frame(table(x)))
temp <- lapply(temp, function(x) arrange(x, desc(Freq)))
lapply(temp, function(x) head(x,10))
```

```
## $`10`
##    x Freq
## 1 US    3
##
## $`11`
##                x Freq
## 1          India    8
## 2             US    4
## 3         Africa    1
## 4 Czechoslovakia    1
##
## $`12`
##              x Freq
## 1        India   26
## 2           US    4
## 3 GreatBritain    3
## 4       Israel    3
## 5   PuertoRico    2
## 6        China    1
## 7      England    1
```

```
## 8   Yugoslavia    1
##
## $`13`
##                x Freq
## 1          India   31
## 2             US    5
## 3         Canada    3
## 4        England    3
## 5         Israel    2
## 6         Africa    1
## 7      Australia    1
## 8   GreatBritain    1
## 9         Mexico    1
## 10     Scotland    1
##
## $`14`
##                 x Freq
## 1           India   54
## 2              US    7
## 3          Israel    6
## 4  UnitedKingdom    5
## 5         England    4
## 6          Africa    2
## 7            Asia    2
## 8       Australia    2
## 9    GreatBritain    2
## 10    Appalachia    1
##
## $`15`
##                 x Freq
## 1           India   28
## 2              US   16
## 3         England   12
## 4  UnitedKingdom   12
## 5          Israel    8
## 6    GreatBritain    6
## 7          Canada    4
## 8          Mexico    3
## 9     PuertoRico    3
## 10          China    2
##
## $`16`
##                 x Freq
## 1          Israel   16
## 2        HongKong   11
## 3              US   11
## 4         England    6
## 5  UnitedKingdom    6
## 6          Canada    4
## 7           India    4
## 8    SouthAfrica    4
## 9           Japan    3
## 10      Scotland    3
##
```

```
## $`17`
##                x Freq
## 1             US  438
## 2         Israel   52
## 3        England   46
## 4         Canada   20
## 5       HongKong   18
## 6      Australia   16
## 7  UnitedKingdom   13
## 8          Wales   10
## 9   GreatBritain    8
## 10         China    7
##
## $`18`
##                x Freq
## 1             US 1151
## 2         Israel  124
## 3         Canada  110
## 4  UnitedKingdom  110
## 5      Australia   80
## 6        England   72
## 7       HongKong   54
## 8         Sweden   36
## 9       Scotland   20
## 10         Wales   19
##
## $`19`
##                x Freq
## 1             US 1645
## 2  UnitedKingdom  281
## 3         Canada  190
## 4      Australia  166
## 5         Israel  123
## 6        England   95
## 7         Sweden   79
## 8       HongKong   63
## 9          China   48
## 10   SouthAfrica   42
##
## $`20`
##                x Freq
## 1             US 2024
## 2      Australia  291
## 3  UnitedKingdom  269
## 4         Canada  210
## 5        England  149
## 6         Israel  114
## 7         Sweden  101
## 8          China   82
## 9       HongKong   68
## 10        Norway   67
```

Methodology

It is easy to explore some of the different fields within the PsychInfo data frame. For example, each record has one or more subject terms (from the article keywords). The total number, unique number, and most frequently occuring key words can be easily computed.

```
MD.df <- filter(pi.df, attributes == "MD")


subject.terms <- stringr::str_split(MD.df$record, pattern = ";")
subject.terms <- unlist(lapply(subject.terms, function(x) gsub(" ", "", x)))
subject.terms.total <- length(unlist(lapply(subject.terms, function(x) gsub(" ", "", x))))
subject.terms.unique <- length(unique(subject.terms))

subject.terms.l <- list(subject.terms.total = subject.terms.total,
                        subject.terms.unique = subject.terms.unique)

most.frequent <- as.data.frame(table(subject.terms))



most.frequent <- arrange(most.frequent, desc(Freq))
most.frequent.t <- head(most.frequent, 50)

print(subject.terms.l)
```

```
$subject.terms.total
[1] 25380

$subject.terms.unique
[1] 21
```

```
print(most.frequent.t)
```

```
                   subject.terms  Freq
1                  EmpiricalStudy 11741
2               QuantitativeStudy  4296
3                QualitativeStudy  3455
4                       Interview  2300
5                LiteratureReview   879
6               LongitudinalStudy   584
7                     FocusGroup   469
8               ClinicalCaseStudy   423
9            NonclinicalCaseStudy   319
10                 FollowupStudy   288
11                    FieldStudy   133
12               SystematicReview   109
13             RetrospectiveStudy   100
14 TreatmentOutcome/ClinicalTrial    84
15               ProspectiveStudy    69
16                   MetaAnalysis    63
17             MathematicalModel    34
```

```
18        ExperimentalReplication    27
19           ScientificSimulation     5
20                   BrainImaging     1
21                      TwinStudy     1
```

Methodology

```
decade <- filter(pi.df, attributes == "YR") %>%
    mutate(year = as.numeric(record)) %>% select(-record, -attributes)

decade$year <- cut(decade$year, breaks = 20, labels = c(1:20))

keywords <- pi.df %>%
        filter(attributes == "MD") %>%
        select(articleID = articleID, keywords = record)

keywords.decade <- keywords %>%
         left_join(decade)
```

```
## Joining by: "articleID"
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
keywords.data.split <- dlply(keywords.decade, .(year))
detach(package:plyr)

terms.f <- function(x){
    split.terms <- stringr::str_split(x[,"keywords"], pattern =";")
    clean.terms <- lapply(split.terms, function(x) gsub(" ", "", x))
    }

keywords.decade <- lapply(keywords.data.split, terms.f)
keywords.decade <- lapply(keywords.decade, unlist)
lapply(keywords.decade, function(x) length(unique(x)))
```

```
## $`4`
## [1] 1
```

```
##
## $`5`
## [1] 1
##
## $`6`
## [1] 1
##
## $`7`
## [1] 1
##
## $`8`
## [1] 1
##
## $`9`
## [1] 1
##
## $`10`
## [1] 5
##
## $`11`
## [1] 6
##
## $`12`
## [1] 6
##
## $`13`
## [1] 5
##
## $`14`
## [1] 8
##
## $`15`
## [1] 10
##
## $`16`
## [1] 8
##
## $`17`
## [1] 16
##
## $`18`
## [1] 16
##
## $`19`
## [1] 18
##
## $`20`
## [1] 21
```

```
temp <- lapply(keywords.decade, function(x) data.frame(table(x)))
temp <- lapply(temp, function(x) arrange(x, desc(Freq)))
lapply(temp, function(x) head(x,10))
```

```
## $`4`
```

```
##           x Freq
## 1 Interview    1
##
## $`5`
##           x Freq
## 1 Interview    5
##
## $`6`
##           x Freq
## 1 Interview    3
##
## $`7`
##           x Freq
## 1 Interview    6
##
## $`8`
##           x Freq
## 1 Interview   10
##
## $`9`
##           x Freq
## 1 Interview    9
##
## $`10`
##                   x Freq
## 1    EmpiricalStudy    9
## 2         Interview    5
## 3 QuantitativeStudy    4
## 4  LiteratureReview    1
## 5  QualitativeStudy    1
##
## $`11`
##                   x Freq
## 1    EmpiricalStudy   20
## 2 QuantitativeStudy    8
## 3 ClinicalCaseStudy    6
## 4         Interview    5
## 5     FollowupStudy    3
## 6  LiteratureReview    2
##
## $`12`
##                   x Freq
## 1 ClinicalCaseStudy   10
## 2  LiteratureReview    8
## 3         Interview    5
## 4    EmpiricalStudy    2
## 5     FollowupStudy    1
## 6 LongitudinalStudy    1
##
## $`13`
##                   x Freq
## 1  LiteratureReview   19
## 2 ClinicalCaseStudy   13
## 3    EmpiricalStudy    7
```

```
## 4     FollowupStudy    7
## 5         Interview    2
##
## $`14`
##                   x Freq
## 1    EmpiricalStudy  535
## 2  LiteratureReview   49
## 3 ClinicalCaseStudy   22
## 4     FollowupStudy    6
## 5         Interview    6
## 6 LongitudinalStudy    1
## 7      MetaAnalysis    1
## 8  SystematicReview    1
##
## $`15`
##                                 x Freq
## 1                 EmpiricalStudy  960
## 2               ClinicalCaseStudy   47
## 3                LiteratureReview   30
## 4                   FollowupStudy   17
## 5                       Interview   10
## 6               LongitudinalStudy    7
## 7         ExperimentalReplication    4
## 8                     MetaAnalysis    3
## 9  TreatmentOutcome/ClinicalTrial    2
## 10           NonclinicalCaseStudy    1
##
## $`16`
##                         x Freq
## 1          EmpiricalStudy 1052
## 2        LiteratureReview   40
## 3           FollowupStudy   24
## 4       ClinicalCaseStudy   21
## 5       LongitudinalStudy   20
## 6               Interview    9
## 7            MetaAnalysis    3
## 8 ExperimentalReplication    1
##
## $`17`
##                                 x Freq
## 1                 EmpiricalStudy 1114
## 2                LiteratureReview   70
## 3               LongitudinalStudy   56
## 4               ClinicalCaseStudy   42
## 5                   FollowupStudy   29
## 6            NonclinicalCaseStudy   26
## 7  TreatmentOutcome/ClinicalTrial   15
## 8                       Interview   10
## 9         ExperimentalReplication    5
## 10               QualitativeStudy    5
##
## $`18`
##                                 x Freq
## 1                 EmpiricalStudy 1757
```

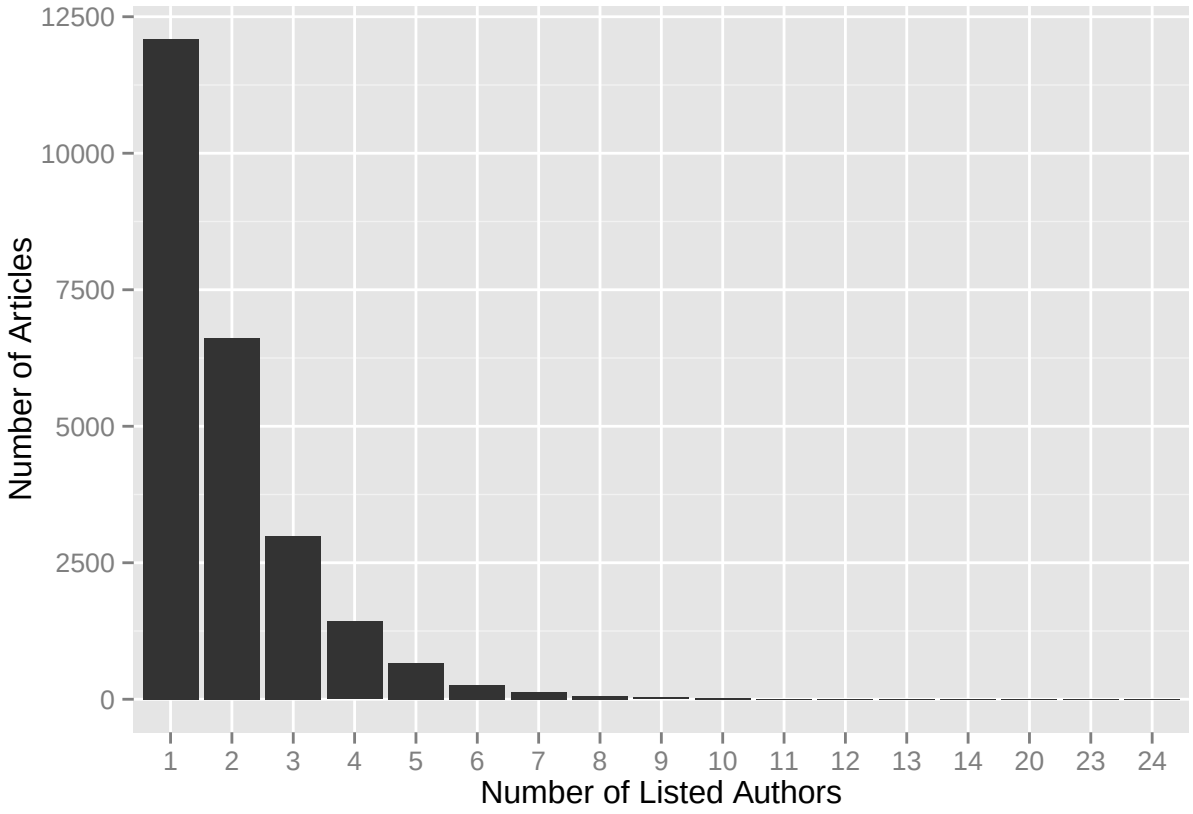
```
## 2               QuantitativeStudy  487
## 3                QualitativeStudy  372
## 4                LiteratureReview  184
## 5             NonclinicalCaseStudy   81
## 6                ClinicalCaseStudy   72
## 7                LongitudinalStudy   72
## 8                    FollowupStudy   40
## 9                       Interview   28
## 10 TreatmentOutcome/ClinicalTrial   20
##
## $`19`
##                        x Freq
## 1        EmpiricalStudy 2682
## 2     QuantitativeStudy 1671
## 3      QualitativeStudy 1220
## 4      LiteratureReview  176
## 5             Interview  158
## 6     LongitudinalStudy  154
## 7     ClinicalCaseStudy   89
## 8  NonclinicalCaseStudy   89
## 9         FollowupStudy   61
## 10           FocusGroup   30
##
## $`20`
##                        x Freq
## 1        EmpiricalStudy 3603
## 2     QuantitativeStudy 2122
## 3             Interview 2028
## 4      QualitativeStudy 1857
## 5            FocusGroup  428
## 6      LiteratureReview  300
## 7     LongitudinalStudy  273
## 8  NonclinicalCaseStudy  122
## 9             FieldStudy  116
## 10    ClinicalCaseStudy  101
```

Number of authors

```
n.authors.article <- pi.df %>%
    filter(attributes == "AU") %>%
    select(id = articleID, author= record) %>%
    mutate(id = as.numeric(id))

n_authors <- n.authors.article %>%
        group_by(id) %>%
        summarise(n = n())

ggplot(n_authors, aes(x = factor(n))) +
    geom_bar() +
    stat_bin(binwidth=1) +
    xlab("Number of Listed Authors") +
    ylab("Number of Articles")
```




```
summary(n_authors$n)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    1.00    2.00    1.94    2.00   24.00
```

Number of authors over time

This figure shows the average number of authors, along with the standard deviation as the ribbon around the average. Note that there is a possible problem in these data, with a single article listing a huge number. That can be corrected at a later time.

```
df.2 <- tbl_df(pi.df)
year <- df.2 %>%
        filter(attributes == "YR") %>%
        select(id = articleID, year = record)

authors <- df.2 %>%
        filter(attributes == "AU") %>%
        select(id = articleID, author = record)

n_authors <- authors %>%
        group_by(id) %>%
        summarise(n=n())

n_authors <- n_authors %>%
        left_join(year) %>%
```

```
        group_by(year) %>%
        summarise(median.n = median(n),
                  average.n = mean(n),
                  min.n = min(n),
                  max.n = max(n),
                  std.dev  = sd(n) )

plot.author.count2 <- ggplot(n_authors, aes(as.numeric(year), y=average.n, group=1)) +
    geom_line(colour="black") +
    geom_ribbon(aes(ymin = average.n-std.dev, ymax=average.n+std.dev), alpha=.2)

head(n_authors, 20)
```

```
Source: local data frame [20 x 6]

   year median.n average.n min.n max.n std.dev
1  1914        1     1.000     1     1      NA
2  1915        1     1.000     1     1      NA
3  1918        1     1.000     1     1      NA
4  1928        1     1.000     1     1  0.0000
5  1929        1     1.000     1     1      NA
6  1930        1     1.000     1     1  0.0000
7  1931        1     1.267     1     2  0.4577
8  1932        1     1.077     1     2  0.2774
9  1933        1     2.600     1    23  5.6543
10 1934        1     1.143     1     2  0.3780
11 1935        1     1.167     1     2  0.4082
12 1936        2     1.750     1     3  0.7071
13 1937        1     1.152     1     4  0.5658
14 1938        1     1.632     1     6  1.3000
15 1939        1     1.070     1     3  0.3082
16 1940        1     1.111     1     3  0.4237
17 1941        1     1.118     1     3  0.4093
18 1942        1     1.139     1     5  0.6825
19 1943        1     1.077     1     4  0.4804
20 1944        1     1.000     1     1  0.0000
```

```
tail(n_authors, 20)
```

```
Source: local data frame [20 x 6]

   year median.n average.n min.n max.n std.dev
1  1995        2     1.921     1     8   1.214
2  1996        2     1.879     1     9   1.185
3  1997        1     1.739     1     6   1.038
4  1998        2     1.934     1     9   1.271
5  1999        1     1.800     1     9   1.185
6  2000        2     1.853     1     8   1.138
7  2001        1     1.842     1     8   1.179
8  2002        2     1.966     1     7   1.155
9  2003        2     1.936     1    20   1.371
10 2004        2     1.928     1     8   1.201
11 2005        2     2.053     1    11   1.388
```

```
12 2006        2     2.122     1    12   1.402
13 2007        2     2.138     1    14   1.488
14 2008        2     2.145     1    12   1.429
15 2009        2     2.246     1    12   1.486
16 2010        2     2.241     1    12   1.488
17 2011        2     2.340     1    13   1.555
18 2012        2     2.305     1    24   1.583
19 2013        2     2.459     1    14   1.660
20 2014        2     2.430     1    12   1.618
```

```
plot.author.count2
```

How Many International Contributors?

This section shows a proof of concept – that is, we can potentially extract all the countries from the author affiliation `AF` tag in the data set. This involves using a set of regular expressions for the extraction. Here I have hard-coded a few countries, but I can obtain a file of all countries and use that to automate the process. We will need to look at the raw data to ensure that the author affiliations have remained in a consistent format throughtout the entirety of the study.

```
df.affiliations <- pi.df %>%
                filter(attributes == "AF")

us.aff <- ifelse(grepl("US", df.affiliations$record, perl=TRUE) == TRUE, "US",
          ifelse(grepl("Canada", df.affiliations$record, perl=TRUE ) == TRUE, "Canada",
          ifelse(grepl("Kong", df.affiliations$record, perl=TRUE ) == TRUE, "Hong Kong",
          ifelse(grepl("China", df.affiliations$record, perl=TRUE) == TRUE, "China",
          ifelse(grepl("Israel", df.affiliations$record, perl=TRUE) == TRUE, "Israel", "Other" )))))
```

```
affiliations <- data.frame(cbind(df.affiliations,us.aff))

ggplot(data=df.affiliations, aes(x = factor(us.aff))) + geom_bar()
```