100 Years of Social Work Research: A Data Science Perspective

Overview of data

Search strategy: January 8, 2014 SO "social work" or SO "social welfare" or SO "human services" or SO "social casework" or SO "social services" or SO "human services"
Limiters - Publication Type: Peer Reviewed Journal; Document Type: Journal Article Search modes - Boolean/Phrase

The search results were exported in a *generic bibliographic format*, which is an unstructured text (*.txt) file.The text file was processed using the `BibWrangleR` function created by the first author.

Initialize OS-X workspace and functions for data wrangling

This section processes raw data. This section of code is executed only one time to transform raw text 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 for Mac OS
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
path <- "/Users/beperron/Git/SocialWorkResearch"

# Wrangle the data with the BWR function suite
#piBWR.f(csv=FALSE, path=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())
setwd("/Users/beperron/Git/SocialWorkResearch")
source("/Users/beperron/Git/BibWrangleR/functions/ggsurv.R")
load("piArticles.R")
library(dplyr)
```

```
library(ggplot2)
library(gridExtra)
library(survival)
library(grid)
library(png)


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

```
[1] 486832      3
```

```
# Inspect variable names 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] 23505
```

Additional cleaning is required. Some

```
pi.df <- pi.df %>% filter(
        record != "Journal of Applied Social Sciences" &
        record != "Early Child Development and Care" &
        record != "The Clinical Supervisor" &
        record != "Children and Youth Services Review" &
        record != "General Hospital Psychiatry" &
        record != "Canadian Journal on Aging" &
        record != "Canadian Journal of Community Mental Health" &
        record != "Behavior Modification" &
        record != "Employee Assistance Quarterly" &
        record != "Journal of Applied Behavioral Science" &
        record != "The Scientific Review of Mental Health Practice: Objective Investigations of Controv
```

What is the overall number and names of journal titles?

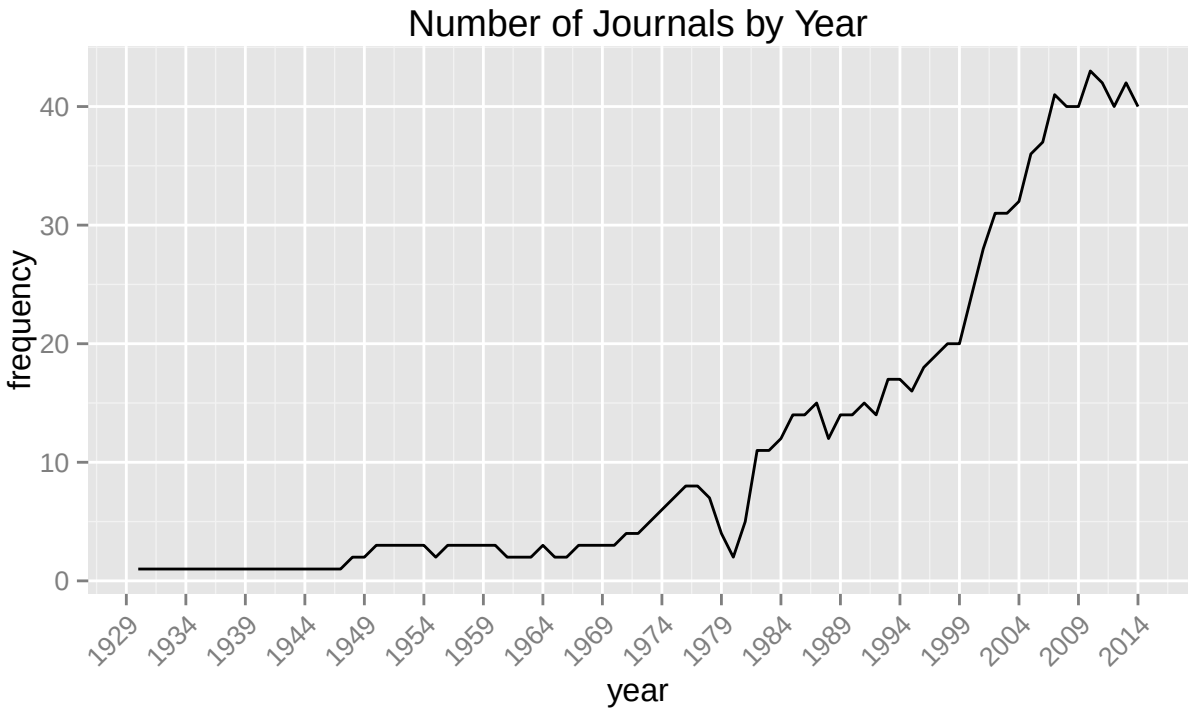
```
unique.titles <- filter(pi.df, attributes == "SO")

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

```
## [1] 55
```

```
# Unique titles
journals.unique <- unique(unique.titles$record)
#write.csv(journals.unique, "journals.csv")
```

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
```

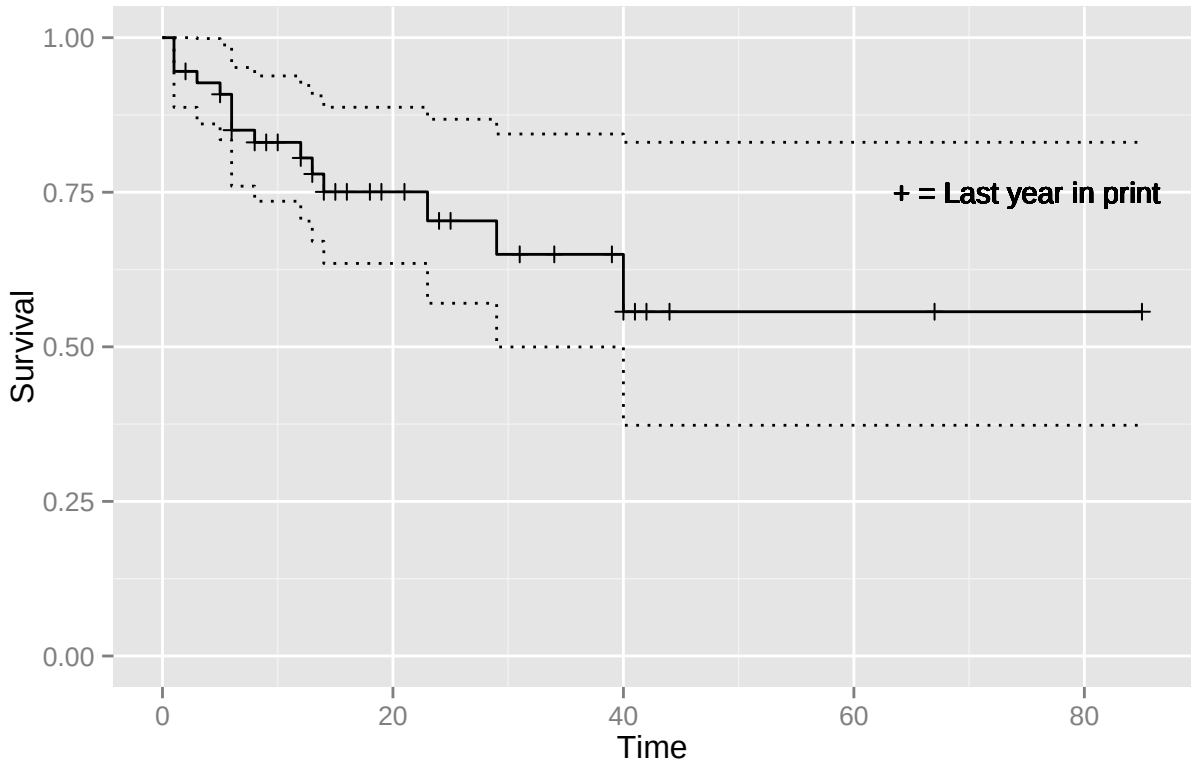
```
## 6                        Social Casework  1950 1989      1095
## 7   Smith College Studies in Social Work  1930 2014      1075
## 8           Clinical Social Work Journal  1973 2014      1074
## 9       Research on Social Work Practice  1991 2014       986
## 10                  Health & Social 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) %>%
       mutate(stop = year.diff, event = ifelse(as.numeric(last) != 2014, 1, 0)) %>%
       select(title, stop, event, as.numeric(first))

survival.journals <- survfit(Surv(longest.running$stop+1, longest.running$event) ~ 1)
median.survival <- data.frame(time = c(12,12), quant = c(.5,0))

head(longest.running)
```

```
## Source: local data frame [6 x 4]
##
##                                    title stop event first
## 1    Smith College Studies in Social Work   84     0  1930
## 2                             Social Work   66     0  1948
## 3          British Journal of Social Work   43     0  1971
## 4            Clinical Social Work Journal   41     0  1973
## 5 Journal of Sociology and Social Welfare   40     0  1974
## 6              Social Work in Health Care   39     0  1975
```

```
ggsurv(survival.journals) +
    #geom_line(data = median.survival, aes(time, quant), linetype="longdash") +
    #annotate("segment", x = 18, xend = 12, y = .12, yend = .15, size = .25, arrow =arrow()) +
    #geom_text(x = 29, y =.12, label = "median survival", size = 4) +
    geom_text(x = 75, y =.75, label = "+ = Last year in print", size = 4) +
    ylim(0,1)
```



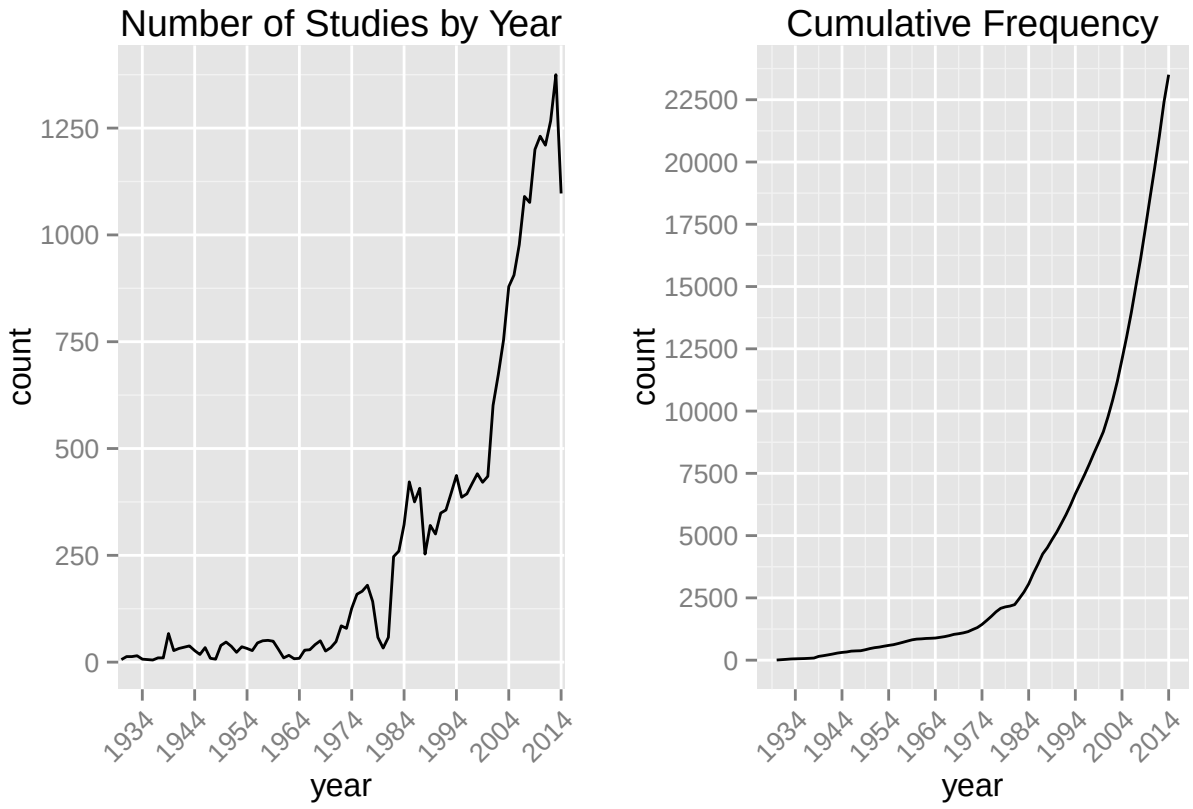
What is the number of articles published per year

```
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

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)
```




```
# Print most recent ten years
tail(df, 10)
```

```
##    years year.count
## 76  2005        906
## 77  2006        977
## 78  2007       1090
## 79  2008       1076
## 80  2009       1200
## 81  2010       1231
## 82  2011       1210
## 83  2012       1267
## 84  2013       1375
## 85  2014       1097
```

What are the topic areas (by Subject Terms)?

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

subject.terms <- stringr::str_split(su.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))

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

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

# Print 25 most commmon terms
head(most.frequent, 25)
```

```
                subject.terms Freq
1              SocialCasework 5703
2               SocialWorkers 2901
3         SocialWorkEducation 1744
4              SocialServices 1184
5                ChildWelfare  826
6               SocialSupport  590
7                  Caregivers  575
8           CommunityServices  569
9                      Family  558
10                ChildAbuse  542
11            MentalDisorders  492
12                 DrugAbuse  490
13                     Aging  477
14                       HIV  473
15       MentalHealthServices  465
16                 FosterCare  462
17            FamilyRelations  461
18               HumanFemales  455
19         HealthCareServices  443
20                MentalHealth  441
21                     Blacks  433
22               Intervention  408
23             CopingBehavior  396
24         GroupPsychotherapy  377
25 PsychotherapeuticProcesses  374
```

What are the topic areas over time (by Subject terms)?

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

```
decade$year <- cut(decade$year, breaks = 10, labels = c(1:10))

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

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

library(plyr)
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)

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))
```

```
## $`1`
##               x Freq
## 1 ChildGuidance    2
##
## $`2`
##               x Freq
## 1        Agency    4
## 2 ChildGuidance    1
##
## $`3`
##               x Freq
## 1        Agency    5
## 2 ChildGuidance    2
##
## $`4`
##                       x Freq
## 1                Agency    9
## 2          ChildGuidance    3
## 3  EmotionalDisturbances    2
## 4             Infidelity    2
## 5          SocialWorkers    2
## 6                Clients    1
## 7        FamilyRelations    1
## 8          FamilyTherapy    1
## 9               Judgment    1
## 10        SocialCasework    1
##
## $`5`
```

```
##                           x Freq
## 1             SocialCasework  147
## 2          CommunityServices   30
## 3                  Treatment   30
## 4              FamilyTherapy   17
## 5            FamilyRelations   15
## 6                 Adjustment   14
## 7                     Family   14
## 8  InterpersonalInteraction   14
## 9       ParentChildRelations   14
## 10       PsychiatricPatients   14
##
## $`6`
##                               x Freq
## 1                SocialCasework  276
## 2                 SocialWorkers  128
## 3              FamilyRelations   55
## 4                FamilyTherapy   53
## 5   PsychotherapeuticProcesses   50
## 6            GroupPsychotherapy   40
## 7           SocialWorkEducation   37
## 8                  HumanFemales   36
## 9  PsychotherapeuticTechniques   36
## 10                  Counseling   35
##
## $`7`
##                       x Freq
## 1        SocialCasework  687
## 2         SocialWorkers  258
## 3       GroupCounseling  154
## 4        SocialServices  136
## 5         SocialSupport  105
## 6         FamilyTherapy   90
## 7   GroupPsychotherapy   87
## 8        CopingBehavior   74
## 9       FamilyRelations   73
## 10 SocialWorkEducation   73
##
## $`8`
##                   x Freq
## 1   SocialCasework  724
## 2    SocialWorkers  284
## 3   SocialServices  170
## 4       ChildAbuse  138
## 5  FamilyRelations  126
## 6    SocialSupport  121
## 7       Caregivers  119
## 8  MentalDisorders  116
## 9             AIDS  105
## 10 GroupCounseling   99
##
## $`9`
##                       x Freq
## 1        SocialCasework 1084
```

```
## 2        SocialWorkers  593
## 3  SocialWorkEducation  307
## 4       SocialServices  295
## 5         ChildWelfare  220
## 6           Caregivers  160
## 7    CommunityServices  145
## 8        SocialSupport  139
## 9           ChildAbuse  137
## 10           DrugAbuse  137
##
## $`10`
##                      x Freq
## 1       SocialCasework 2784
## 2        SocialWorkers 1626
## 3  SocialWorkEducation 1236
## 4       SocialServices  583
## 5         ChildWelfare  529
## 6         Intervention  344
## 7                Aging  331
## 8               Family  316
## 9         MentalHealth  292
## 10                HIV  291
```

What are the most frequent topic areas (by author specified keywords)?

```
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)

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

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

# Print summary statistics
print(subject.terms.l)
```

```
$subject.terms.total
[1] 99607

$subject.terms.unique
[1] 45826
```

```r
# Print 25 most frequent
head(most.frequent, 25)
```

```
             subject.terms Freq
1               socialwork 1777
2            socialworkers 1762
3      socialworkeducation  787
4       socialworkpractice  552
5           socialservices  343
6       socialworkstudents  315
7             mentalhealth  312
8                 children  277
9             childwelfare  258
10                    HIV  232
11          socialsupport  227
12 CHILDHOODANDADOLESCENCE  217
13             riskfactors  200
14            spirituality  200
15          decisionmaking  189
16              fostercare  188
17           socialjustice  186
18                  aging  179
19        domesticviolence  176
20            intervention  176
21             adolescents  168
22              depression  161
23                 poverty  159
24         childprotection  158
25  conferencepresentation  158
```

Most Frequent Author Keywords

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

decade$year <- cut(decade$year, breaks = 10, labels = c(1:10))

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

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

library(plyr)
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)

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))
```

```
## $`1`
##                                        x Freq
## 1                 CHILDHOODANDADOLESCENCE   51
## 2                                  CHILD   33
## 3          SOCIALFUNCTIONSOFTHEINDIVIDUAL   19
## 4              NERVOUSANDMENTALDISORDERS   13
## 5                            SOCIALABILITY   10
## 6  GENERALSOCIALPROCESSES(INCL.ESTHETICS)    9
## 7                                    WORK    9
## 8                     FUNCTIONALDISORDERS    8
## 9                             PERSONALITY    8
## 10                            ADJUSTMENT    7
##
## $`2`
##                                        x Freq
## 1                 CHILDHOODANDADOLESCENCE  164
## 2                     FUNCTIONALDISORDERS   68
## 3                                GUIDANCE   66
## 4                 CHILD(IV.MALADJUSTMENT   62
## 5                                THERAPY)   62
## 6                                   CHILD   58
## 7  GENERALSOCIALPROCESSES(INCL.ESTHETICS)   44
## 8          CHILD(MALADJUSTMENTANDTHERAPY)   42
## 9                                  CLINIC   33
## 10                            ADJUSTMENT   32
##
## $`3`
##                        x Freq
## 1             SOCIALWORK   76
## 2                   CASE   53
## 3             TECHNIQUES   53
## 4          CHILDGUIDANCE   52
## 5            METHODOLOGY   52
## 6       TREATMENTMETHODS   34
## 7         SOCIALCASEWORK   33
## 8             COUNSELING   29
## 9                 FAMILY   27
## 10 CHILDHOOD&ADOLESCENCE   26
##
## $`4`
##                   x Freq
## 1     SOCIALWELFARE   71
## 2       METHODOLOGY   70
## 3        TECHNIQUES   70
```

```
## 4            FAMILY   33
## 5  TREATMENTMETHODS   29
## 6     CHILDGUIDANCE   28
## 7     PSYCHOTHERAPY   27
## 8         CHILDHOOD   23
## 9  BEHAVIORPROBLEMS   20
## 10   SOCIALCASEWORK   18
##
## $`5`
##                                    x Freq
## 1                      socialworkers   10
## 2                            clients    6
## 3                      socialcasework    6
## 4                         socialwork    5
## 5  CHILDHOOD/EMOTIONALDISTURBANCESIN    4
## 6                      SOCIALCASEWORK    4
## 7                          casereport    3
## 8                 COUNSELING&GUIDANCE    3
## 9                        grouptherapy    3
## 10                           REVIEWS    3
##
## $`6`
##                                x Freq
## 1                socialworkers   32
## 2              literaturereview   10
## 3                     socialwork    8
## 4            socialworkstudents    8
## 5                          aged    7
## 6                socialcasework    7
## 7                      children    6
## 8    implicationsforsocialwork    5
## 9  implicationsforsocialworkers    5
## 10                familytherapy    4
##
## $`7`
##                                x Freq
## 1                socialworkers  137
## 2              literaturereview   50
## 3    implicationsforsocialwork   43
## 4       conferencepresentation   41
## 5                       elderly   35
## 6                      children   27
## 7                    casereport   22
## 8  implicationsforsocialworkers   20
## 9                        Israel   20
## 10                         aged   19
##
## $`8`
##                                x Freq
## 1                socialworkers  123
## 2       conferencepresentation  115
## 3              literaturereview   65
## 4    implicationsforsocialwork   57
## 5                        Israel   54
```

```
## 6                       England   45
## 7                    casereport   39
## 8                        Canada   31
## 9                       elderly   29
## 10 implicationsforsocialworkers   29
##
## $`9`
##                       x Freq
## 1           socialwork  397
## 2         socialworkers  390
## 3    socialworkpractice  149
## 4   socialworkeducation  129
## 5              children   85
## 6          mentalhealth   74
## 7        socialservices   68
## 8         socialsupport   65
## 9    socialworkstudents   65
## 10           depression   59
##
## $`10`
##                       x Freq
## 1           socialwork 1353
## 2         socialworkers 1069
## 3   socialworkeducation  652
## 4    socialworkpractice  384
## 5        socialservices  273
## 6          mentalhealth  232
## 7    socialworkstudents  216
## 8          childwelfare  205
## 9                   HIV  191
## 10          riskfactors  166
```

Location of Studies

```
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))

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

print(subject.terms.l)
```

```
$subject.terms.total
```

```
[1] 11181

$subject.terms.unique
[1] 207
```

```
print(location)
```

```
                      subject.terms Freq
1                                US 5457
2                     UnitedKingdom  690
3                            Canada  570
4                         Australia  563
5                            Israel  450
6                           England  389
7                            Sweden  231
8                          HongKong  220
9                             China  160
10                       SouthAfrica  115
11                            Norway  102
12                        NewZealand   98
13                             India   92
14                          Scotland   91
15                           Ireland   87
16                             Wales   84
17                           Finland   74
18                           Germany   73
19                       Netherlands   62
20                    NorthernIreland   52
21                           Denmark   51
22                             Spain   49
23                       GreatBritain   44
24                             Italy   44
25                            Taiwan   43
26                             Japan   42
27                           Belgium   41
28                         Singapore   41
29                            Africa   36
30                            Mexico   36
31                            Europe   33
32                            France   31
33                            Greece   28
34                            Russia   28
35                           Romania   25
36                             Ghana   24
37                             Korea   24
38                            Uganda   23
39                            Brazil   22
40                           Hungary   22
41                          Portugal   22
42                          Thailand   22
43                           Nigeria   18
44                       PuertoRico   18
45                       SouthKorea   18
46                           Vietnam   18
```

47	Asia	17
48	Austria	17
49	Switzerland	17
50	Kenya	16
51	NorthAmerica	16
52	Croatia	15
53	Poland	15
54	Turkey	15
55	Chile	14
56	Ethiopia	14
57	Botswana	13
58	Iran	13
59	Zambia	13
60	Malaysia	12
61	Slovenia	11
62	CzechRepublic	10
63	Zimbabwe	10
64	Bangladesh	9
65	Estonia	9
66	Iceland	9
67	Tanzania	9
68	Argentina	8
69	Bulgaria	8
70	Georgia	8
71	Guatemala	8
72	Jordan	8
73	Lithuania	8
74	Luxembourg	8
75	Nepal	8
76	Palestine	8
77	Peru	8
78	Albania	7
79	Caribbean	7
80	Colombia	7
81	Cyprus	7
82	Egypt	7
83	ElSalvador	7
84	Pakistan	7
85	Philippines	7
86	Rwanda	7
87	Ukraine	7
88	USSR	7
89	Latvia	6
90	Lebanon	6
91	SaudiArabia	6
92	SriLanka	6
93	TrinidadandTobago	6
94	UnitedArabEmirates	6
95	Afghanistan	5
96	Cambodia	5
97	Cuba	5
98	DominicanRepublic	5
99	Ecuador	5
100	Indonesia	5

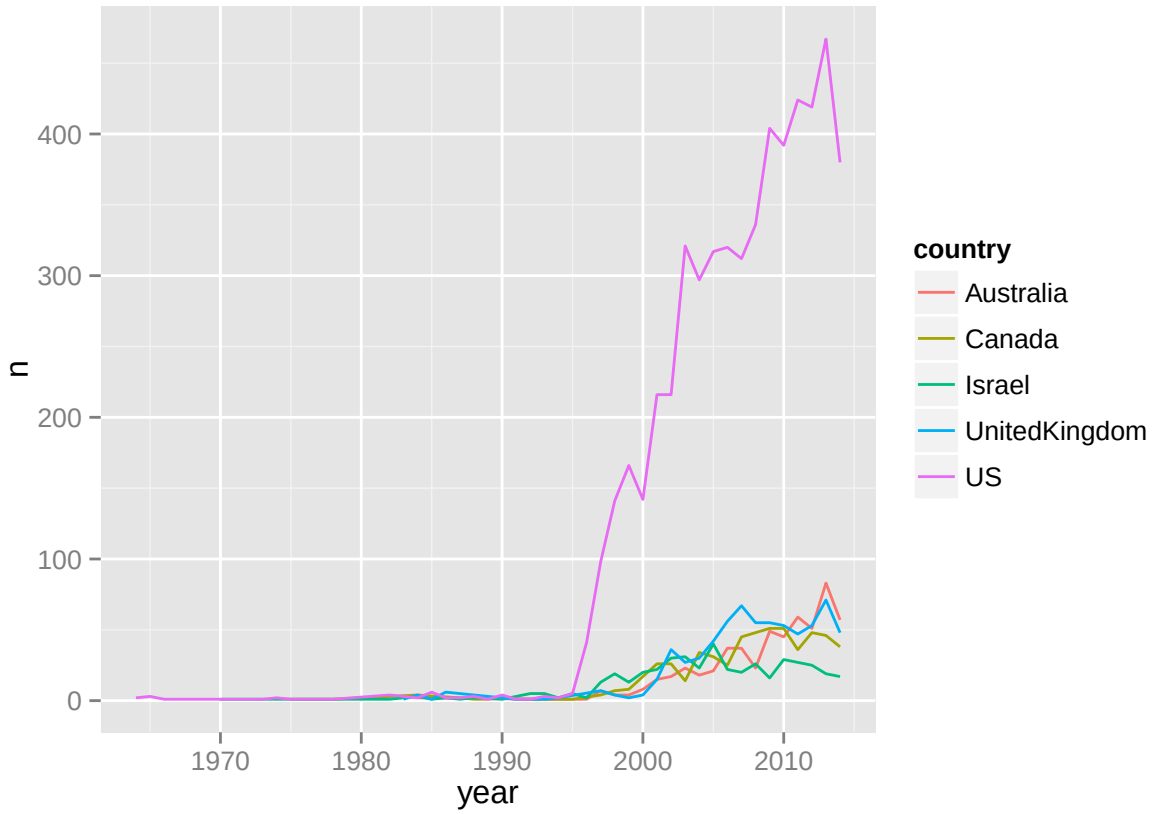
101	Kuwait	5
102	Moldova	5
103	Mongolia	5
104	Slovakia	5
105	Somalia	5
106	Azerbaijan	4
107	Barbados	4
108	Bosnia-Herzegovina	4
109	Cameroon	4
110	DemocraticRepublicofCongo	4
111	Kazakhstan	4
112	Lesotho	4
113	Nicaragua	4
114	SierraLeone	4
115	SouthAmerica	4
116	Tajikistan	4
117	Bolivia	3
118	CentralAmerica	3
119	CostaRica	3
120	Guyana	3
121	Haiti	3
122	Honduras	3
123	Iraq	3
124	Jamaica	3
125	Kyrgyzstan	3
126	Liberia	3
127	Malawi	3
128	Malta	3
129	MarshallIslands	3
130	Mauritius	3
131	RepublicofSerbia	3
132	Yugoslavia	3
133	Appalachia	2
134	Armenia	2
135	Belarus	2
136	Bermuda	2
137	Bhutan	2
138	Czechoslovakia	2
139	EasternEurope	2
140	Fiji	2
141	Gambia	2
142	Macedonia	2
143	MiddleEast	2
144	Morocco	2
145	Mozambique	2
146	Myanmar	2
147	NewCaledonia	2
148	Oceania/PacificIslands	2
149	Palau	2
150	Panama	2
151	PapuaNewGuinea	2
152	Paraguay	2
153	Samoa	2
154	Scandinavia	2

155	Swaziland	2
156	Tonga	2
157	Tuvalu	2
158	Uruguay	2
159	Uzbekistan	2
160	WesternEurope	2
161	Yemen	2
162	Algeria	1
163	Angola	1
164	Bahamas	1
165	Bahrain	1
166	BalticStates	1
167	Brunei	1
168	Burundi	1
169	ChannelIslands	1
170	CommonwealthofIndependentStates	1
171	Comoros	1
172	CookIslands	1
173	Eritrea	1
174	FrenchPolynesia	1
175	Gabon	1
176	Grenada	1
177	Guam	1
178	Guinea	1
179	IvoryCoast	1
180	Kiribati	1
181	Laos	1
182	LatinAmerica	1
183	Liechtenstein	1
184	Macau	1
185	Madagascar	1
186	Maldives	1
187	Micronesia(FederatedStatesof)	1
188	Montenegro	1
189	Namibia	1
190	Nauru	1
191	Niue	1
192	NorthKorea	1
193	Oman	1
194	Qatar	1
195	RepublicofCongo	1
196	Senegal	1
197	SerbiaandMontenegro	1
198	SlovakRepublic	1
199	SolomonIslands	1
200	StKitts	1
201	Sudan	1
202	Togo	1
203	USVirginIslands	1
204	Vanuatu	1
205	Venezuela	1
206	WestBank	1
207	WestIndies	1

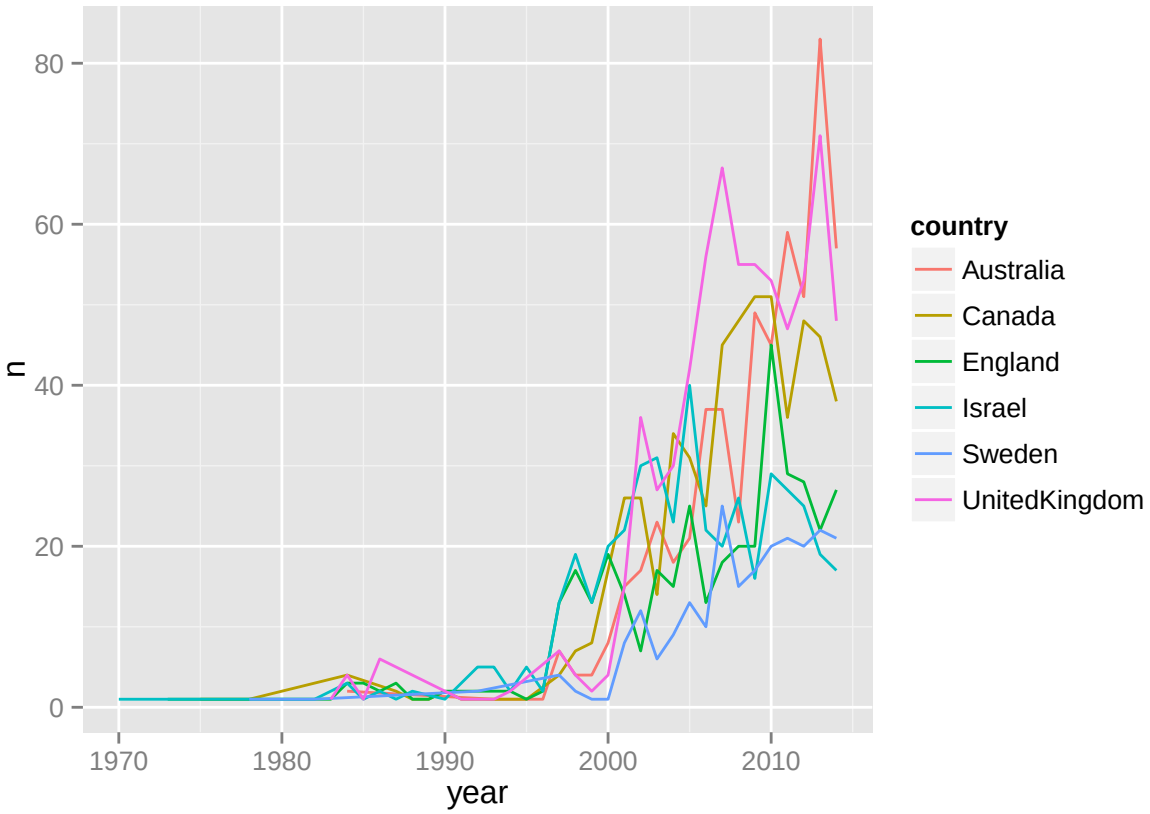
```
#img2 <- readPNG("/Users/beperron/Git/SocialWorkResearch/Chloro.png")
#grid.raster(img2)
```

Location of studies over time

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

location.temp <- pi.df %>%
        filter(attributes == "LO") %>%
        select(articleID = articleID, country = record) %>%
        mutate(country = gsub(" ", "", country))

location.year <- location.temp %>%
         left_join(year) %>%
         group_by(country, year) %>%
         summarise(n = n()) %>%
         arrange(country, year, n)

top.locations <- location %>%
                 select(subject.terms) %>%
                 rename(country = subject.terms) %>%
                 slice(1:5)

location.year.top <- location.year %>% filter(country %in% top.locations$country)
ggplot(location.year.top, aes(x = year, y = n, colour = country )) + geom_line()
```




```
top.locations.no.us <- location %>%
                select(subject.terms) %>%
                rename(country = subject.terms) %>%
                slice(2:7)

location.year.top.no.us <- location.year %>% filter(country %in% top.locations.no.us$country)

ggplot(location.year.top.no.us, aes(x = year, y = n, colour = country )) + geom_line()
```



Countries not represented in social work research

This section is currently not functional

```
#Tableau.countries <- read.csv("/Users/beperron/Git/SocialWorkResearch/SupportingDocs/Tableau\ Countrie
```

Methodology

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

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

methodology.table <- as.data.frame(table(methodology.terms))
methodology.table <- arrange(methodology.table, desc(Freq))
```

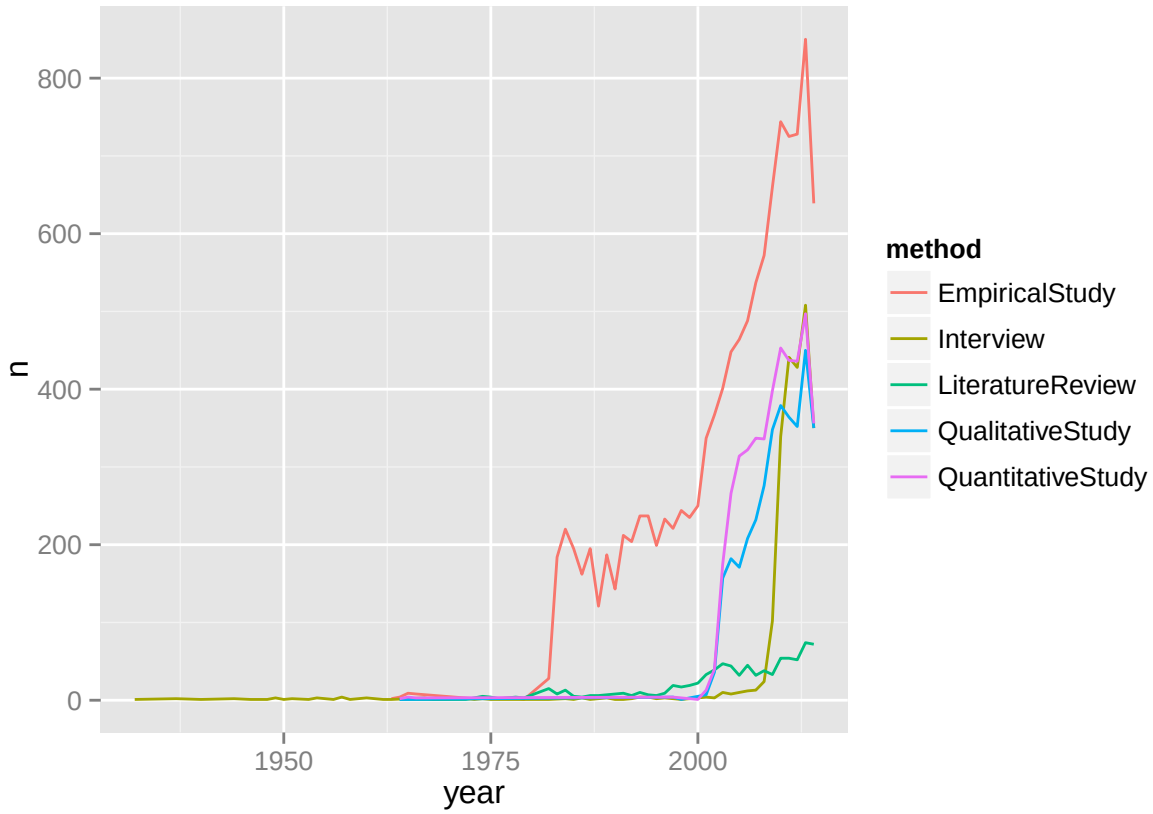
Methodology over time

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

```
methodology <- pi.df %>%
        filter(attributes == "MD") %>%
        select(articleID = articleID, method = record) %>%
        mutate(method = gsub(" ", "", method))

methodology.year <- methodology %>%
         left_join(year) %>%
         group_by(method, year) %>%
         summarise(n = n()) %>%
         arrange(method, year, n)
```

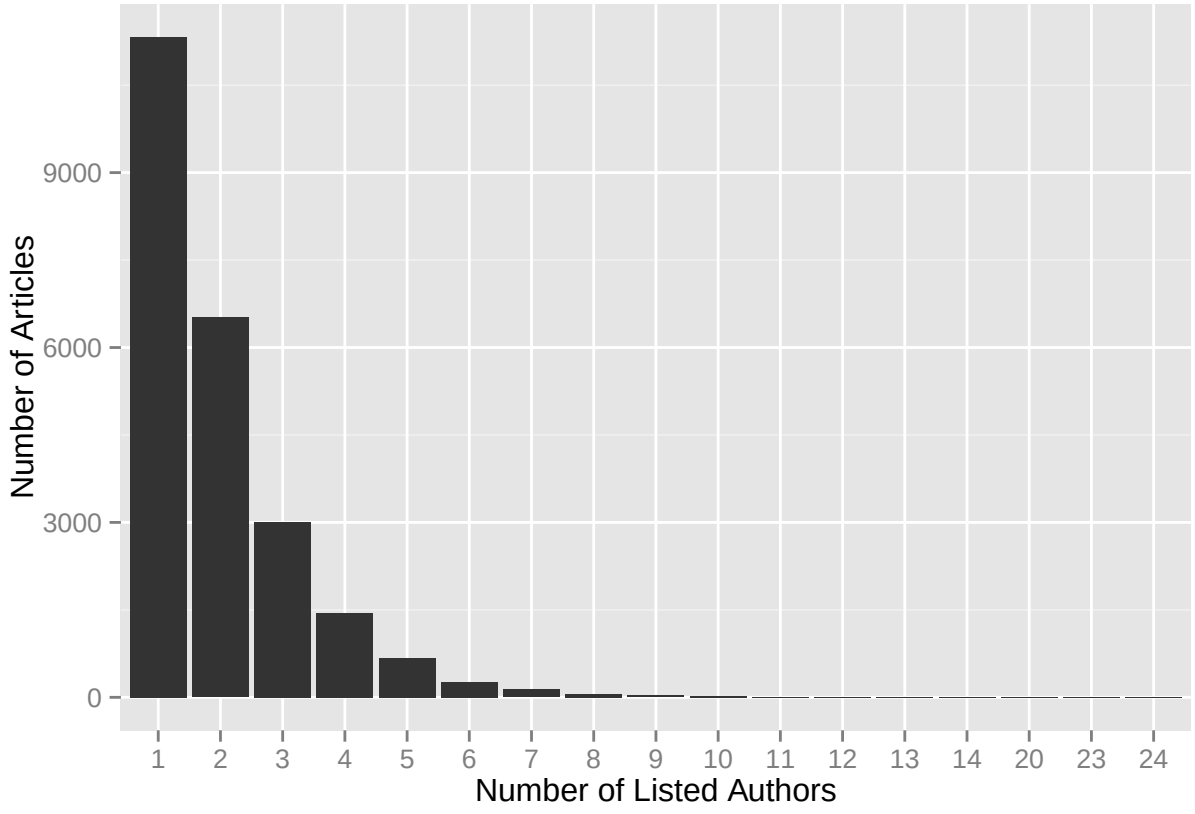
```
## Joining by: "articleID"
```

```
methodology.year.top <- slice(methodology.table, 1:5) %>%
         rename(method = methodology.terms) %>%
         mutate(method = gsub(" ", "", method))

methodology.year.reduced <- methodology.year %>%
        filter(method %in% methodology.year.top$method)

ggplot(methodology.year.reduced, aes(x = year, y = n, colour = method)) +
        geom_line()
```



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    1.97    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  1930        1     1.000     1     1  0.0000
2  1931        1     1.308     1     2  0.4804
3  1932        1     1.077     1     2  0.2774
4  1933        1     2.600     1    23  5.6543
5  1934        1     1.143     1     2  0.3780
6  1935        1     1.167     1     2  0.4082
7  1936        1     1.600     1     3  0.8944
8  1937        1     1.100     1     2  0.3162
9  1938        1     2.000     1     6  1.6997
10 1939        1     1.045     1     3  0.2715
11 1940        1     1.111     1     3  0.4237
12 1941        1     1.094     1     3  0.3902
13 1942        1     1.143     1     5  0.6921
14 1943        1     1.079     1     4  0.4867
15 1944        1     1.000     1     1  0.0000
16 1945        1     1.056     1     2  0.2357
17 1946        1     1.029     1     2  0.1715
18 1947        1     1.222     1     2  0.4410
19 1948        1     1.000     1     1  0.0000
20 1949        1     1.026     1     2  0.1601
```

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

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

   year median.n average.n min.n max.n std.dev
1  1995      2.0     1.904     1     8   1.218
2  1996      1.5     1.865     1     9   1.176
3  1997      1.0     1.746     1     6   1.049
4  1998      2.0     1.982     1     9   1.309
5  1999      1.0     1.805     1     9   1.193
6  2000      2.0     1.867     1     8   1.144
7  2001      1.0     1.834     1     8   1.171
8  2002      2.0     1.961     1     7   1.141
9  2003      2.0     1.928     1    20   1.362
10 2004      2.0     1.928     1     8   1.199
11 2005      2.0     2.063     1    11   1.383
12 2006      2.0     2.127     1    12   1.401
13 2007      2.0     2.126     1    14   1.479
14 2008      2.0     2.157     1    12   1.446
15 2009      2.0     2.245     1    12   1.486
16 2010      2.0     2.234     1    12   1.484
17 2011      2.0     2.334     1    13   1.545
18 2012      2.0     2.293     1    24   1.573
19 2013      2.0     2.460     1    14   1.657
20 2014      2.0     2.416     1    12   1.607
```

```
plot.author.count2
```