

# DA5020.Practium1.Luo.Rmd

Xin Lin, Xinxin Luo, William Morrison

```
library(tidyverse)
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.2.1    v purrr  0.3.3
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   1.0.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
```

```
## -- Conflicts -----
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(ggplot2)
library(dplyr)
library(readr)
library(XML)
library(RCurl)
```

```
##
```

```
## Attaching package: 'RCurl'
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##     complete
```

```
library(lubridate)
```

```
##
```

```
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
```

```
##
```

```
##     date
```

```
library(knitr)
```

## Question 1

Load the XML file into a browser or text editing tool and inspect it. Explore the data set as you see fit and that allows you to get a sense of the data and get comfortable with it.

```

#Loading the dataset
xmldata<-getURL("https://da5020.weebly.com/uploads/8/6/5/9/8659576/pubmedsample.jun18.xml")
xmldf<-xmlToDataFrame(xmldata)

# Convert data into tree
parsed_data<-xmlTreeParse(xmldata, useInternalNodes = TRUE)
medRoot<-xmlRoot(parsed_data)

```

## Question 2

Load the data into R and create two linked tibbles: one for publication and one for journal. Use ISSN as the key to link them. Only load the following information into the publication tibble: PMID (primary key for publication), ISSN and publication year (foreign key for journal), date completed (as one date field), date revised (as one date field), number of authors (derived/calculated field from the authors), publication type, title of article. Load this information into the journal tibble: ISSN (primary key), medium (from CitedMedium attribute), publication year (primary key), publication season, language, and journal title. In cases where there are multiple languages for a publication, pick the first language. Same for publication type: pick the first one. The primary key for journal is (ISSN, publication year). Also, exclude any journals that do not have an ISSN as the primary key cannot be empty.

```

# Parsing information for tibbles
# PMID number
parsePMID <- function(rNode) {
  pmID <- xmlValue(rNode[["MedlineCitation"]][["PMID"]][["text"]])
  return(pmID)
}

# ISSN number
parseISSN <- function(rNode) {
  issn <- xmlValue(rNode[["MedlineCitation"]][["Article"]][["Journal"]][["ISSN"]][["text"]])
  return(issn)
}

# Publication year
parsePubYear <- function(rNode) {
  pubYear <- xmlValue(rNode[["MedlineCitation"]][["Article"]][["Journal"]][["JournalIssue"]][["PubDate"]])
  return(pubYear)
}

# Date completed
parseCompDate <- function(rNode) {
  compDate <- xmlValue(rNode[["MedlineCitation"]][["DateCompleted"]])
  return(compDate)
}

# Date revised
parseRevDate <- function(rNode) {
  revDate <- xmlValue(rNode[["MedlineCitation"]][["DateRevised"]])
  return(revDate)
}

# Number of authors

```

```

parseNumAuth <- function(rNode) {
  authors <- xmlSize(rNode[["MedlineCitation"]][["Article"]][["AuthorList"]])
  numAuthors <- as.numeric(authors)
  return(numAuthors)
}

# Publication Type
parsePubType <- function(rNode) {
  pubType <- xmlValue(rNode[["MedlineCitation"]][["Article"]][["PublicationTypeList"]][["PublicationType"]])
  return(pubType)
}

# Article Title
parseArtTitle <- function(rNode) {
  title <- xmlValue(rNode[["MedlineCitation"]][["Article"]][["ArticleTitle"]][["text"]])
  return(title)
}

# Cited Medium
Medium <- xpathSApply(parsed_data, "//PubmedArticle/MedlineCitation", function(x) {
  if (xpathSApply(x, "boolean(./Article/Journal/JournalIssue)") {
    xpathSApply(x, "./Article/Journal/JournalIssue", xmlGetAttr, "CitedMedium")
  }
  else {
    NA
  }
})

# Publication Season
parsePubSeason <- function(rNode) {
  pubSeason <- xmlValue(rNode[["MedlineCitation"]][["Article"]][["Journal"]][["JournalIssue"]][["PubDate"]])
  return(pubSeason)
}

# Language
parseLang <- function(rNode) {
  lang <- xmlValue(rNode[["MedlineCitation"]][["Article"]][["Language"]])
  return(lang)
}

# Journal Title
parseJourTitle <- function(rNode) {
  title <- xmlValue(rNode[["MedlineCitation"]][["Article"]][["Journal"]][["Title"]][["text"]])
  return(title)
}

# Construct tibbles####
# Initialize tibbles
publication <- tibble(pmID = character(),
                      issn = character(),
                      pubYear = character(),
                      dateComp = date(),
                      dateRevi = date(),

```

```

        authors = numeric(),
        pubType = character(),
        title = character())
journal <- tibble(issn = character(),
                 medium = character(),
                 pubYear = date(),
                 pubSeason = character(),
                 lang = character(),
                 jourTitle = character())

# Loop through the tree to get information
numNodes <- xmlSize(medRoot)

for (i in 1:numNodes)
{
  # turn off parse warnings
  options(warn = -1)

  # work on the "ith" listing
  currListing <- medRoot[[i]]

  # parse auction ID and add as PK for listing tibble
  pk <- parsePMID(currListing)
  # ensure that PK is unique, i.e., is not already in listing tibble
  if (nrow(publication[publication$pmID == pk,]) == 1)
  {
    # it already exists, so somehow the listing is already there so skip everything below
    # and continue with the next iteration of the loop
    print (paste("pmID already exists:", pk), quote = FALSE)
    next
  }

  # add data to publication tibble
  publication[i,]$pmID <- pk
  publication[i,]$issn <- parseISSN(currListing)
  publication[i,]$pubYear <- parsePubYear(currListing)
  publication[i,]$dateComp <- parseCompDate(currListing)
  publication[i,]$dateRevi <- parseRevDate(currListing)
  publication[i,]$authors <- parseNumAuth(currListing)
  publication[i,]$pubType <- parsePubType(currListing)
  publication[i,]$title <- parseArtTitle(currListing)

  # add data to journal tibble
  journal[i,]$issn <- parseISSN(currListing)
  journal[i,]$medium <- Medium[i]
  journal[i,]$pubYear <- parsePubYear(currListing)
  journal[i,]$pubSeason <- parsePubSeason(currListing)
  journal[i,]$lang <- parseLang(currListing)
  journal[i,]$jourTitle <- parseJourTitle(currListing)
}

# Change the class to "Dates"
publication$dateComp <- ymd(publication$dateComp, tz = "UTC")

```

```

publication$dateRevi <- ymd(publication$dateRevi, tz = "UTC")

# Extract valid records
journal <- filter(journal, issn != "NA", pubYear != "NA")
publication <- filter(publication, issn != "NA", pubYear != "NA")
journal <- unique(journal)

# Display the publication and journal tibbles
kable(publication)

```

pmID	issn	pubYear	dateComp	dateRevi	authors	pubType	title
973217	0095-3814	1976	1976-12-03	2002-11-13	1	Journal Article	Hospital del
1669026	0377-8231	1991	1993-11-15	2008-11-21	0	Congresses	[150th Anni
1875346	0022-2623	1991	1991-09-25	2013-11-21	10	Journal Article	3-Hydroxy-3
3549656	0021-8820	1987	1987-05-15	2014-11-20	6	Comparative Study	Semisynthet
8119288	0014-2956	1994	1994-04-01	2016-10-17	2	Journal Article	Purification
8655018	0017-0011	1996	1996-07-30	2017-03-03	2	English Abstract	[Effect of fet
8863847	0026-895X	1996	1996-11-25	2014-11-20	5	Journal Article	Mechanism o
9110943	1059-2725	1996	1996-10-23	2011-11-17	4	Journal Article	Conservative
8941094	0009-7322	1996	1997-01-02	2016-11-24	11	Journal Article	Assessment
9394824	0014-2980	1997	1998-01-02	2006-11-15	3	Journal Article	Hypermutat
9482442	0140-6736	1998	1998-03-18	2016-11-24	4	Case Reports	A woman w
9505772	0007-0912	1998	1998-03-24	2004-11-17	8	Clinical Trial	Myocardial
9575322	0002-838X	1998	1998-05-12	2007-11-15	2	Journal Article	Lumbar spin
9627170	1435-2443	1998	1998-08-25	2016-10-20	3	Journal Article	Outcome of
9634358	0022-1899	1998	1998-06-18	2004-03-31	7	Retraction of Publication	Retraction:
9626910	0047-1828	1998	1998-08-12	2015-11-19	25	Journal Article	Hepatitis C
9861576	0268-1315	1998	1999-03-29	2013-11-21	5	Clinical Trial	Cardiac side
9929727	0300-2896	1998	1999-04-28	2017-11-16	3	Letter	[Tobacco co
9885794	0893-133X	1999	1999-03-29	2015-03-11	3	Journal Article	Antipsychot
10078868	0041-0101	1999	1999-06-09	2005-11-17	0	Bibliography	Bibliography
10083987	0832-610X	1999	1999-05-11	2004-11-17	1	Comment	Complete ai
10101342	0733-8627	1999	1999-04-15	2016-11-24	2	Journal Article	Evaluation o
10168751	0168-8510	1997	1997-08-28	2006-11-15	2	Comparative Study	Health techn
10097079	0027-8424	1999	1999-05-12	2016-10-19	6	Comparative Study	Thermal ada
10192114	0031-7144	1999	1999-06-01	2007-01-29	2	Journal Article	Antimicrobi
10192115	0031-7144	1999	1999-06-01	2013-11-21	4	Journal Article	Different kin
10331748	0031-9023	1999	1999-06-01	2009-11-11	1	Editorial	Looking for
10331749	0031-9023	1999	1999-06-01	2009-11-11	4	Journal Article	Effects of sic
10389168	0869-8139	1999	1999-07-15	2016-10-20	4	English Abstract	[Ethanol mo
10540283	0950-382X	1999	1999-12-17	2006-11-15	8	Journal Article	Transcriptio
10854512	1432-2218	2000	2000-06-29	2004-11-17	2	Case Reports	Inflammator
10737756	0022-2623	2000	2000-04-13	2013-11-21	25	Journal Article	Phosphoryla
10612833	1098-1004	2000	2000-01-20	2012-07-11	7	Corrected and Republished Article	Erratum: an
10972993	0899-1987	2000	2000-09-26	2008-11-21	6	Journal Article	Altered exp
11025314	0108-2701	2000	2001-01-05	2006-11-15	5	Journal Article	Multicentre
11034741	1469-493X	2000	2001-08-02	2013-06-28	2	Journal Article	Parent-train
11056631	0031-9007	2000	2016-03-08	2000-12-01	3	Journal Article	Dislocated e
11238657	0022-1767	2001	2001-05-17	2013-11-21	8	Journal Article	Histamine in
11279676	1469-493X	2001	2001-12-07	2014-11-20	2	Journal Article	Elective ver
11428848	0195-668X	2001	2001-09-27	2007-11-15	8	Guideline	Indications f
11431089	0924-7947	2001	2005-06-08	2004-04-20	9	Journal Article	An investiga
11406024	1469-493X	2001	2002-02-28	2013-06-28	2	Journal Article	Parent-train

pmID	issn	pubYear	dateComp	dateRevi	authors	pubType	title
11441930	0284-186X	2001	2001-07-19	2015-11-19	4	Journal Article	Assessment
11442735	0303-6979	2001	2001-10-04	2013-11-21	10	Clinical Trial	Utilisation of
11237011	0028-0836	2001	2001-03-22	2016-10-25	257	Journal Article	Initial sequen
11488864	0905-7161	2001	2001-10-25	2013-11-21	6	Comparative Study	Early endoss
11473127	0021-9258	2001	2001-11-01	2013-11-21	9	Journal Article	Evidence tha
11488868	0905-7161	2001	2001-10-25	2012-11-15	5	Case Reports	Histology of
11520209	0022-2623	2001	2001-09-27	2013-11-21	9	Journal Article	Novel azo de
11525160	0250-5525	2000	2001-08-30	2004-11-17	0	Congresses	32nd Annual
11524736	1098-1004	2001	2001-12-12	2016-11-24	4	Journal Article	Detection of
11537092	0033-183X	1991	1995-06-16	2017-08-16	3	Journal Article	Cytoplasmic
11540070	0735-2689	1987	1995-12-11	2013-11-21	2	Journal Article	Calcium me
11543891	0198-8859	2001	2001-09-27	2011-11-17	9	Comparative Study	The origin o
11567133	0036-8075	2001	2001-10-04	2007-03-19	2	Journal Article	Changes in s
11580607	0031-9007	2001	2001-10-25	2006-11-15	4	Journal Article	Maximal he
11583040	0084-6597	2000	2001-10-04	2006-11-15	2	Journal Article	Understandi
11600210	0198-8859	2001	2001-10-25	2001-10-15	2	Retraction of Publication	Editorial. A
11600211	0198-8859	2001	2001-10-25	2004-11-17	1	Comment	Letter from
11612527	0032-4728	1990	1991-03-18	2014-11-20	1	Historical Article	On the origi
11620107	0511-0726	1984	1988-11-16	2016-10-21	2	Historical Article	Demographi
11675721	0024-2160	1974	1976-07-01	2016-11-23	1	Biography	Thomas Dov
11694165	0098-7484	2001	2001-12-28	2016-10-17	1	Journal Article	msJAMA: E
11704686	0028-4793	2001	2001-12-28	2011-11-17	1	Journal Article	Recognition
11731716	0031-7012	2002	2002-02-28	2005-11-16	1	Journal Article	Progress in
11748851	1098-1004	2001	2002-03-07	2016-11-24	4	Journal Article	Erratum: D
11748856	1098-1004	2001	2002-03-07	2016-11-24	5	Journal Article	Eight novel
11831940	0003-9950	2002	2002-02-25	2015-04-17	4	Comment	Republicati
11838066	0023-7205	2002	2002-02-25	2014-11-20	1	Biography	[Unequal dis
11858625	0002-838X	2002	2002-03-06	2004-11-17	0	Journal Article	Information
11869993	1052-1577	2002	2002-03-22	2016-10-20	1	Journal Article	By the way
11876201	1069-9422	1998	2002-03-22	2016-11-24	5	Journal Article	Consumer a
11885531	0028-2200	1992	2002-04-08	2006-11-15	2	English Abstract	[Microbiolog
11892742	0012-1606	2002	2002-03-18	2009-07-23	2	Corrected and Republished Article	The major s
11916658	1070-910X	2002	2002-04-15	2016-11-24	1	Journal Article	By the way,
11953811	1432-2218	2002	2002-05-24	2017-10-10	1	Editorial	Disruptive v
11966386	0098-7484	2002	2002-05-02	2016-10-17	13	Consensus Development Conference	The 2001 Be
12038483	0273-1177	2000	2002-07-26	2013-11-21	1	Journal Article	Planetary p
12101218	0021-9258	2002	2002-08-12	2006-11-15	4	Journal Article	Interaction l
12159900	0362-4331	2002	2002-08-22	2013-06-07	1	Interview	Weighing m
12174865	1145-0762	2000	2002-08-21	2004-11-17	2	Journal Article	Physicians' a
12145319	0021-9258	2002	2002-09-16	2013-11-21	4	Journal Article	Amisyn, a n
12192682	0018-9251	1999	2002-09-09	2006-11-15	2	Journal Article	Intelligent c
12179763	1226-0282	1990	1991-12-03	2015-09-01	3	Journal Article	Recent chan
12211241	1095-9203	2002	2002-09-13	2007-03-19	1	Comment	Mycobacteri
12211266	1054-6863	2002	2002-09-25	2004-11-17	1	Journal Article	Public polic
12219757	1225-505X	2001	2002-09-13	2016-11-24	1	Biography	Developmen
12227380	0269-8897	2002	2002-09-13	2004-11-17	1	Biography	Paracelsus,
12230384	1539-3704	2002	2002-09-25	2005-11-17	0	Journal Article	Summaries f
12230355	1539-3704	2002	2002-09-25	2005-11-17	1	Guideline	Screening fo
12349809	0251-7329	2000	2000-12-06	2007-11-15	1	Journal Article	Gender and
12369570	8750-7587	2002	2003-02-21	2013-09-26	1	Comment	Human unil
12501715	0559-7765	1998	2003-01-08	2005-11-16	1	Journal Article	[Mechanism
12599353	0025-4282	1994	2003-03-10	2007-11-15	1	Journal Article	The new Un

pmID	issn	pubYear	dateComp	dateRevi	authors	pubType	title
12599354	0025-4282	1994	2003-03-10	2004-11-17	1	Journal Article	The right to
11909345	0031-9007	2002	2002-05-16	2003-11-03	621	Journal Article	Measuremen
12742516	1879-0712	2003	2009-02-27	2013-11-21	4	Journal Article	S-15176 inh
12742519	1879-0712	2003	2009-02-27	2014-11-20	5	Journal Article	Risperidone
12742518	1879-0712	2003	2009-02-27	2014-11-20	5	Journal Article	Behavioral e
13634534	0014-2565	1958	2000-07-01	2013-11-21	7	Journal Article	[Physiopath
14177620	0016-5662	1963	1996-12-01	2006-11-15	4	Journal Article	[JEJUNAL
14337379	0002-9378	1965	1996-12-01	2017-09-08	3	Journal Article	SEROTONI
14594616	1087-2108	2003	2004-02-02	2004-11-17	2	Case Reports	Epidermal m
14668029	1607-8454	2003	2015-02-20	2003-12-11	4	Journal Article	Clopidogrel:
14668030	1607-8454	2003	2015-02-20	2003-12-11	3	Journal Article	Platelet cou
14695699	1226-0479	2003	2005-04-12	2013-11-21	9	English Abstract	[The signific
14729922	1362-4962	2004	2004-02-11	2014-06-10	1	Journal Article	Local homol
14749521	1539-6150	2004	2004-02-23	2013-11-21	1	News	Hampering
14744982	1362-4962	2004	2004-02-02	2014-06-10	3	Journal Article	Superior 5' l
14796701	0028-0836	1950	2004-02-15	2015-11-19	2	Journal Article	Reduction b
14991249	0364-2348	2004	2005-03-15	2017-09-23	3	Case Reports	Spontaneous
15117668	0022-2895	1990	2004-05-10	2004-04-30	1	Journal Article	The First C
15110832	0968-0896	2004	2004-12-10	2006-11-15	6	Journal Article	Aminyl and
15206831	0896-4289	2003	2004-07-29	2007-11-14	4	Journal Article	Warm partn
15228064	1080-6032	2004	2004-08-06	2013-11-21	5	Case Reports	Acute coron
15278624	0913-8668	1991	2005-02-07	2017-08-12	3	Journal Article	Hypoxic ven
15284444	0027-8424	2004	2004-09-29	2017-02-19	2	Journal Article	Tubular pre
15611660	1551-4005	2005	2006-03-23	2017-11-16	2	Journal Article	Altered epig
15611661	1551-4005	2005	2006-03-23	2009-11-19	5	Journal Article	TrkAIII. A r
15611667	1551-4005	2005	2006-03-23	2013-11-21	9	Journal Article	Replication
15739502	1000-8713	1997	2010-04-15	2016-10-18	2	English Abstract	[The determ
15810377	1000-0593	1997	2010-05-21	2005-04-06	2	English Abstract	[Study on cl
16776646	0248-4900	2006	2006-08-09	2016-10-20	1	Journal Article	Books for fr
15968009	1539-3704	2005	2005-06-27	2014-07-29	23	Clinical Trial	The effect o
16779244	1942-597X	2005	2007-02-15	2016-10-25	3	Journal Article	MeSH Spell
16888359	1064-3745	2006	2006-09-28	2014-09-09	2	Journal Article	Mining micr
16923641	0803-9488	2006	2007-01-25	2006-08-22	1	Journal Article	The use of E
17059514	0269-2813	2006	2007-05-07	2006-10-24	14	Journal Article	Effectiveness
16949478	1558-3597	2006	2006-09-21	2006-09-04	46	Journal Article	ACC/AHA/
17712873	1432-2218	2007	2007-08-22	2007-10-01	12	Journal Article	Crura ultras
17713168	1359-6535	2007	2007-10-19	2007-11-30	12	Journal Article	Declining pr
17823161	1756-1833	2007	2007-09-17	2013-11-21	1	News	Agency war
17926191	1042-8194	2007	2008-04-09	2013-11-21	4	Evaluation Studies	Advanced ag
18311089	1091-0220	2008	2016-04-23	2016-09-12	0	Journal Article	Cancer deat
18243949	0278-0062	1985	2012-10-02	2008-02-04	2	Journal Article	Electrophor
18964660	0039-9140	1988	2012-10-02	2008-10-30	3	Journal Article	Multiparam
18941263	1833-3575	2008	2015-07-23	2008-10-22	2	Journal Article	Issues in the
19771122	0146-9592	1990	2009-12-11	2009-09-22	7	Journal Article	Ultrashort-l
21002435	0002-9955	1946	2010-10-28	2014-08-12	0	Journal Article	BRUTALIT
18719013	1756-1833	2008	2008-09-02	2017-09-22	10	Journal Article	Health relat
23674598	1477-9129	2013	2013-07-16	2017-09-22	2	Journal Article	The neural c
23326248	1612-3174	2013	2013-06-28	2015-02-19	2	Case Reports	Transcathet
23807877	1662-4548	2013	2013-07-01	2017-02-20	2	Journal Article	Improved bl
23685357	1460-2075	2013	2013-08-12	2017-09-22	18	Journal Article	Eps8 contro
23852273	2157-3999	2013	2013-07-15	2017-02-20	4	Journal Article	Twitter as a
25143782	1749-7922	2014	2014-08-21	2014-08-25	2	Journal Article	Reviewer ac

pmID	issn	pubYear	dateComp	dateRevi	authors	pubType	title
24090994	1873-6971	2013	2014-06-02	2013-11-01	7	Journal Article	Modulation
25031417	1529-2401	2014	2014-09-05	2014-11-20	5	Journal Article	A new pathv
27687974	1941-0611	2017	NA	2017-01-03	1	Journal Article	SAR11 Bact
27687975	2045-2322	2016	NA	2017-02-20	9	Journal Article	Novel roles f
27763809	2164-554X	2017	2017-10-24	2017-10-24	5	Comparative Study	Do Australia
29049350	1932-6203	2017	2017-11-07	2017-11-07	14	Journal Article	An integrati
23203861	0167-9228	2007	NA	2012-12-03	0	Journal Article	NA
23203860	0167-9228	2007	NA	2012-12-03	0	Journal Article	NA
28647898	1439-0973	2017	NA	2017-11-21	10	Journal Article	Respiratory
28726616	1080-6059	2017	NA	2017-11-11	20	Journal Article	Molecular C
23203892	0167-9228	2010	NA	2012-12-03	0	Journal Article	NA
23203893	0167-9228	2010	NA	2012-12-03	0	Journal Article	NA
23203895	0167-9228	2010	NA	2012-12-03	0	Journal Article	NA
23203894	0167-9228	2010	NA	2012-12-03	0	Journal Article	NA
23203896	0167-9228	2010	NA	2012-12-03	0	Journal Article	NA

`kable(journal)`

issn	medium	pubYear	pubSeason	lang	jourTitle
0095-3814	Print	1976	Fall	eng	Topics in health care financing
0377-8231	Print	1991	NA	fre	Bulletin et memoires de l'Academie royale de medecine de Belgique
0022-2623	Print	1991	NA	eng	Journal of medicinal chemistry
0021-8820	Print	1987	NA	eng	The Journal of antibiotics
0014-2956	Print	1994	NA	eng	European journal of biochemistry
0017-0011	Print	1996	NA	pol	Ginekologia polska
0026-895X	Print	1996	NA	eng	Molecular pharmacology
1059-2725	Print	1996	NA	eng	The Online journal of current clinical trials
0009-7322	Print	1996	NA	eng	Circulation
0014-2980	Print	1997	NA	eng	European journal of immunology
0140-6736	Print	1998	NA	eng	Lancet (London, England)
0007-0912	Print	1998	NA	eng	British journal of anaesthesia
0002-838X	Print	1998	NA	eng	American family physician
1435-2443	Print	1998	NA	eng	Langenbeck's archives of surgery
0022-1899	Print	1998	NA	eng	The Journal of infectious diseases
0047-1828	Print	1998	NA	eng	Japanese circulation journal
0268-1315	Print	1998	NA	eng	International clinical psychopharmacology
0300-2896	Print	1998	NA	spa	Archivos de bronconeumologia
0893-133X	Print	1999	NA	eng	Neuropsychopharmacology : official publication of the American Col
0041-0101	Print	1999	NA	eng	Toxicon : official journal of the International Society on Toxinology
0832-610X	Print	1999	NA	eng	Canadian journal of anaesthesia = Journal canadien d'anesthesie
0733-8627	Print	1999	NA	eng	Emergency medicine clinics of North America
0168-8510	Print	1997	NA	eng	Health policy (Amsterdam, Netherlands)
0027-8424	Print	1999	NA	eng	Proceedings of the National Academy of Sciences of the United State
0031-7144	Print	1999	NA	eng	Die Pharmazie
0031-9023	Print	1999	NA	eng	Physical therapy
0869-8139	Print	1999	NA	rus	Rossiiskii fiziologicheskii zhurnal imeni I.M. Sechenova
0950-382X	Print	1999	NA	eng	Molecular microbiology
1432-2218	Internet	2000	NA	eng	Surgical endoscopy
0022-2623	Print	2000	NA	eng	Journal of medicinal chemistry
1098-1004	Internet	2000	NA	eng	Human mutation



issn	medium	pubYear	pubSeason	lang	jourTitle
0899-1987	Print	2000	NA	eng	Molecular carcinogenesis
0108-2701	Print	2000	NA	eng	Acta crystallographica. Section C, Crystal structure communications
1469-493X	Internet	2000	NA	eng	The Cochrane database of systematic reviews
0031-9007	Print	2000	NA	eng	Physical review letters
0022-1767	Print	2001	NA	eng	Journal of immunology (Baltimore, Md. : 1950)
1469-493X	Internet	2001	NA	eng	The Cochrane database of systematic reviews
0195-668X	Print	2001	NA	eng	European heart journal
0924-7947	Print	2001	NA	eng	Archives of gerontology and geriatrics. Supplement
0284-186X	Print	2001	NA	eng	Acta oncologica (Stockholm, Sweden)
0303-6979	Print	2001	NA	eng	Journal of clinical periodontology
0028-0836	Print	2001	NA	eng	Nature
0905-7161	Print	2001	NA	eng	Clinical oral implants research
0021-9258	Print	2001	NA	eng	The Journal of biological chemistry
0022-2623	Print	2001	NA	eng	Journal of medicinal chemistry
0250-5525	Print	2000	NA	eng	Schweizerische medizinische Wochenschrift. Supplementum
1098-1004	Internet	2001	NA	eng	Human mutation
0033-183X	Print	1991	NA	eng	Protoplasma
0735-2689	Print	1987	NA	eng	Critical reviews in plant sciences
0198-8859	Print	2001	NA	eng	Human immunology
0036-8075	Print	2001	NA	eng	Science (New York, N.Y.)
0031-9007	Print	2001	NA	eng	Physical review letters
0084-6597	Print	2000	NA	eng	Annual review of earth and planetary sciences
0032-4728	Print	1990	NA	eng	Population studies
0511-0726	Print	1984	NA	eng	A.A.G. bijdragen
0024-2160	Print	1974	NA	eng	The Library
0098-7484	Print	2001	NA	eng	JAMA
0028-4793	Print	2001	NA	eng	The New England journal of medicine
0031-7012	Print	2002	NA	eng	Pharmacology
0003-9950	Print	2002	NA	eng	Archives of ophthalmology (Chicago, Ill. : 1960)
0023-7205	Print	2002	NA	swe	Lakartidningen
0002-838X	Print	2002	NA	eng	American family physician
1052-1577	Print	2002	NA	eng	Harvard health letter
1069-9422	Print	1998	NA	eng	Life support & biosphere science : international journal of earth space
0028-2200	Print	1992	NA	dut	Nederlands tijdschrift voor tandheelkunde
0012-1606	Print	2002	NA	eng	Developmental biology
1070-910X	Print	2002	NA	eng	Harvard women's health watch
1432-2218	Internet	2002	NA	eng	Surgical endoscopy
0098-7484	Print	2002	NA	eng	JAMA
0273-1177	Print	2000	NA	eng	Advances in space research : the official journal of the Committee on
0021-9258	Print	2002	NA	eng	The Journal of biological chemistry
0362-4331	Print	2002	NA	eng	The New York times on the Web
1145-0762	Print	2000	NA	eng	Journal international de bioethique = International journal of bioeth
0018-9251	Print	1999	NA	eng	IEEE transactions on aerospace and electronic systems
1226-0282	Print	1990	NA	eng	Pogon sahoe nonjip = Journal of population, health, and social welfa
1095-9203	Internet	2002	NA	eng	Science (New York, N.Y.)
1054-6863	Print	2002	NA	eng	Kennedy Institute of Ethics journal
1225-505X	Print	2001	NA	eng	Ui sahak
0269-8897	Print	2002	NA	eng	Science in context
1539-3704	Internet	2002	NA	eng	Annals of internal medicine
0251-7329	Print	2000	NA	eng	UN chronicle
8750-7587	Print	2002	NA	eng	Journal of applied physiology (Bethesda, Md. : 1985)
0559-7765	Print	1998	NA	chi	Sheng li ke xue jin zhan [Progress in physiology]

issn	medium	pubYear	pubSeason	lang	jourTitle
0025-4282	Print	1994	NA	eng	Maryland law review (Baltimore, Md. : 1936)
0031-9007	Print	2002	NA	eng	Physical review letters
1879-0712	Internet	2003	NA	eng	European journal of pharmacology
0014-2565	Print	1958	NA	spa	Revista clinica espanola
0016-5662	Print	1963	NA	ita	Gazzetta internazionale di medicina e chirurgia
0002-9378	Print	1965	NA	eng	American journal of obstetrics and gynecology
1087-2108	Internet	2003	NA	eng	Dermatology online journal
1607-8454	Internet	2003	NA	eng	Hematology (Amsterdam, Netherlands)
1226-0479	Print	2003	NA	kor	Taehan Kan Hakhoe chi = The Korean journal of hepatology
1362-4962	Internet	2004	NA	eng	Nucleic acids research
1539-6150	Internet	2004	NA	eng	Science of aging knowledge environment : SAGE KE
0028-0836	Print	1950	NA	eng	Nature
0364-2348	Print	2004	NA	eng	Skeletal radiology
0022-2895	Print	1990	NA	eng	Journal of motor behavior
0968-0896	Print	2004	NA	eng	Bioorganic & medicinal chemistry
0896-4289	Print	2003	Fall	eng	Behavioral medicine (Washington, D.C.)
1080-6032	Print	2004	Summer	eng	Wilderness & environmental medicine
0913-8668	Print	1991	NA	eng	Journal of anesthesia
0027-8424	Print	2004	NA	eng	Proceedings of the National Academy of Sciences of the United States of America
1551-4005	Internet	2005	NA	eng	Cell cycle (Georgetown, Tex.)
1000-8713	Print	1997	NA	chi	Se pu = Chinese journal of chromatography
1000-0593	Print	1997	NA	chi	Guang pu xue yu guang pu fen xi = Guang pu
0248-4900	Print	2006	NA	eng	Biology of the cell
1539-3704	Internet	2005	NA	eng	Annals of internal medicine
1942-597X	Internet	2005	NA	eng	AMIA ... Annual Symposium proceedings. AMIA Symposium
1064-3745	Print	2006	NA	eng	Methods in molecular biology (Clifton, N.J.)
0803-9488	Print	2006	NA	eng	Nordic journal of psychiatry
0269-2813	Print	2006	NA	eng	Alimentary pharmacology & therapeutics
1558-3597	Internet	2006	NA	eng	Journal of the American College of Cardiology
1432-2218	Internet	2007	NA	eng	Surgical endoscopy
1359-6535	Print	2007	NA	eng	Antiviral therapy
1756-1833	Internet	2007	NA	eng	BMJ (Clinical research ed.)
1042-8194	Print	2007	NA	eng	Leukemia & lymphoma
1091-0220	Print	2008	NA	eng	Mayo Clinic women's healthsource
0278-0062	Print	1985	NA	eng	IEEE transactions on medical imaging
0039-9140	Print	1988	NA	eng	Talanta
1833-3575	Internet	2008	NA	eng	The HIM journal
0146-9592	Print	1990	NA	eng	Optics letters
0002-9955	Print	1946	NA	eng	Journal of the American Medical Association
1756-1833	Internet	2008	NA	eng	BMJ (Clinical research ed.)
1477-9129	Internet	2013	NA	eng	Development (Cambridge, England)
1612-3174	Internet	2013	NA	eng	German medical science : GMS e-journal
1662-4548	Print	2013	NA	eng	Frontiers in neuroscience
1460-2075	Internet	2013	NA	eng	The EMBO journal
2157-3999	Internet	2013	NA	eng	PLoS currents
1749-7922	Print	2014	NA	eng	World journal of emergency surgery : WJES
1873-6971	Internet	2013	NA	eng	Fitoterapia
1529-2401	Internet	2014	NA	eng	The Journal of neuroscience : the official journal of the Society for Neuroscience
1941-0611	Internet	2017	NA	eng	Annual review of marine science
2045-2322	Internet	2016	NA	eng	Scientific reports
2164-554X	Internet	2017	NA	eng	Human vaccines & immunotherapeutics
1932-6203	Internet	2017	NA	eng	PloS one

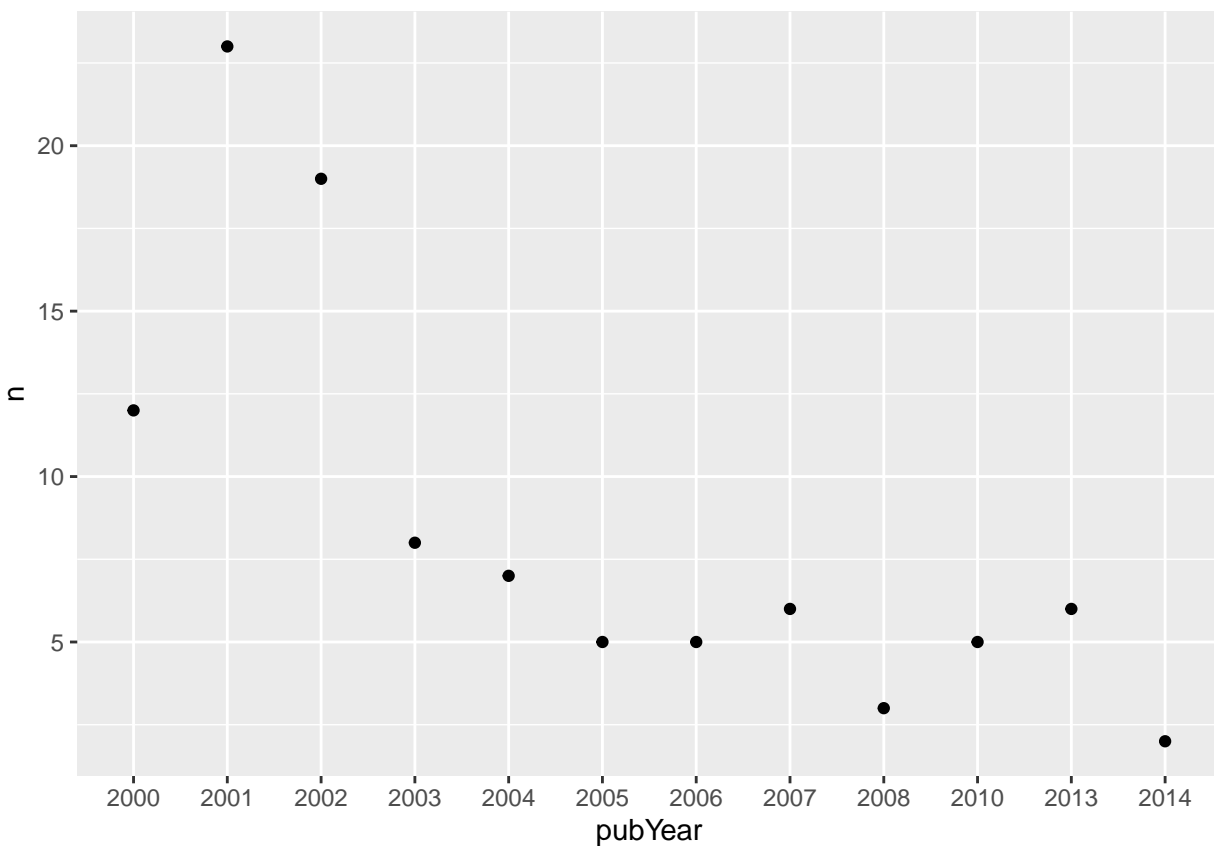
issn	medium	pubYear	pubSeason	lang	jourTitle
0167-9228	Print	2007	NA	dut	Tijdschrift voor gerontologie en geriatrie
1439-0973	Internet	2017	NA	eng	Infection
1080-6059	Internet	2017	NA	eng	Emerging infectious diseases
0167-9228	Print	2010	NA	dut	Tijdschrift voor gerontologie en geriatrie

### Question 3

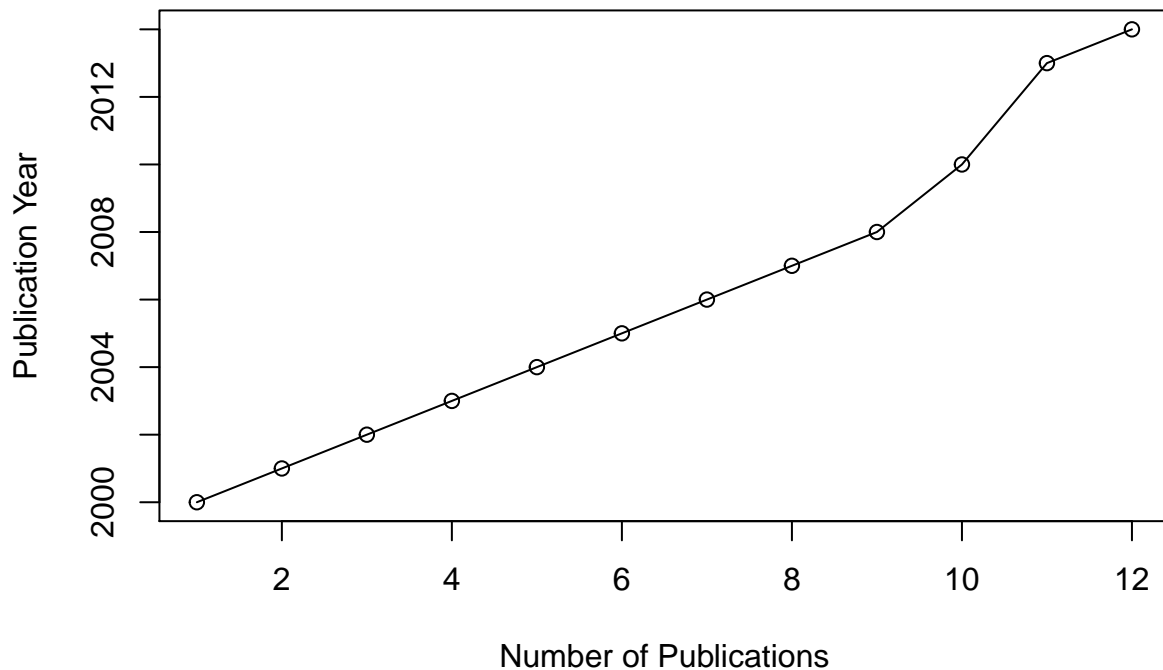
Create a line graph of the number of publications per year from 2000 to 2015.

```
# find number of publications per year from 2000 to 2015
pubsPerYear <- publication %>%
  filter(pubYear >= 2000, pubYear <= 2015) %>%
  count(pubYear)

# Plot the results
ggplot(data = pubsPerYear, aes(x = pubYear, y = n)) +
  geom_point()
```



```
plot(pubsPerYear$pubYear, type = "o", xlab = "Number of Publications", ylab = "Publication Year")
```



#### Question 4

Find the articles that had fewer than three authors and list the article, journal, and publication date.

```
# find articles with 1-3 authors
threeOrLessAuthors <- publication %>%
  inner_join(journal, by = c("issn", "pubYear")) %>%
  filter(authors < 4, authors > 0) %>%
  select(title, jourTitle, dateComp)

# find articles with 0 author
zeroAuthors <- publication %>%
  inner_join(journal, by = c("issn", "pubYear")) %>%
  filter(authors == 0) %>%
  select(title, jourTitle, dateComp)

# print the results
kable(threeOrLessAuthors)
```

---

title

Hospital debt management and cost reimbursement.

Purification and characterisation of a water-soluble ferrochelatase from *Bacillus subtilis*.

[Effect of fetal and neonatal growth on the occurrence of some diseases in adults].

Hypermutation, diversity and dissemination of human intestinal lamina propria plasma cells.

---

title

Lumbar spine stenosis: a common cause of back and leg pain.  
Outcome of patients with sepsis and septic shock after ICU treatment.  
[Tobacco control in children, adolescents and young people: knowledge, prevention and action].  
Antipsychotic potential of CCK-based treatments: an assessment using the prepulse inhibition model of psychosis.  
Complete airway obstruction.  
Evaluation of the patient with extremity trauma: an evidence based approach.  
Health technology assessment: decentralized and fragmented in the US compared to other countries.  
Antimicrobial activity of some Nepalese medicinal plants.  
Looking for the Forrest.  
Inflammatory fibroid polyp of the duodenum.  
Parent-training programmes for improving maternal psychosocial health.  
Dislocated epitaxial islands.  
Elective versus selective caesarean section for delivery of the small baby.  
Parent-training programmes for improving maternal psychosocial health.  
Cytoplasmic calcium levels in protoplasts from the cap and elongation zone of maize roots.  
Calcium messenger system in plants.  
Changes in seismic anisotropy after volcanic eruptions: evidence from Mount Ruapehu.  
Understanding oblique impacts from experiments, observations, and modeling.  
Editorial. Anthropology and genetic markers.  
Letter from the ASHI president and council.  
On the origins of the United States Government's international population policy.  
Demographic transition in the Netherlands: a statistical analysis of regional differences in the level and development of the  
Thomas Dover's "Ancient physician's legacy".  
msJAMA: Breast reconstruction: one woman's choice.  
Recognition and management of anthrax—an update.  
Progress in the search for ideal drugs.  
[Unequal distribution of health resources. Research is not supported with consideration to the global significance of the disease].  
By the way doctor... I'm 72 and was recently told that I am losing a bit of my sight because of macular degeneration. My  
[Microbiological diagnostics in periodontal treatment].  
The major subacrosomal occupant of bull spermatozoa is a novel histone H2B variant associated with the forming acrosome.  
By the way, doctor. A few months ago, I was surprised to read that a new study has raised a question about the value of so  
Disruptive visions: surgeon responsibility during the era of change.  
Planetary protection issues for Mars sample acquisition flight projects.  
Weighing medical ethics for many years to come: a conversation with Harold Shapiro. Interview by Howard Markel.  
Physicians' attitudes towards medical ethics issues in Turkey.  
Intelligent control of a planning system for astronaut training.  
Recent changes in the population control policy and its future directions in Korea.  
Mycobacterium leprae and demyelination.  
Public policy and the sale of human organs.  
Development of neurophysiology in the early twentieth century: Charles Scott Sherrington and The Integrative action of the  
Paracelsus, Paracelsianism, and the secularization of the worldview.  
Screening for osteoporosis in postmenopausal women: recommendations and rationale.  
Gender and globalization. A century in retrospect.  
Human unilateral lower limb suspension as a model for spaceflight effects on skeletal muscle.  
[Mechanism of bone mineral loss in microgravity].  
The new Uniform Health Care Decisions Act: paving a health care decisions superhighway?  
The right to refuse life-sustaining medical treatment: national trend and recent changes in Maryland law.  
SEROTONIN CONTENT OF HUMAN PLACENTA AND FETUS DURING PREGNANCY.  
Epidermal nevus.  
Platelet counts and interleukin-6 (IL-6) promoter polymorphism in patients with Gaucher disease.  
Local homology recognition and distance measures in linear time using compressed amino acid alphabets.  
Hampering a heartbreaker. Antibiotic might stem injury from heart attack.

---

title

Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter.  
Reduction by molecular hydrogen of acetoacetate to butyrate by butyric acid bacteria.  
Spontaneous resolution of solitary osteochondroma in the young adult.  
The First Conference on Motor Control in Down Syndrome.  
Hypoxic ventilatory response in cats lightly anesthetized with ketamine: effects of halothane and sevoflurane in low concentrations.  
Tubular precipitation and redox gradients on a bubbling template.  
Altered epigenetic patterning leading to replicative senescence and reduced longevity. A role of a novel SNF2 factor, PASG.  
[The determination of olaquinox in feeds by HPLC].  
[Study on cleavage mechanisms of dimer (t-Bu)<sub>2</sub>NO in several solvent phases].  
Books for free? How can this be? - A PubMed resource you may be overlooking.  
MeSH Speller + askMEDLINE: auto-completes MeSH terms then searches MEDLINE/PubMed via free-text, natural language.  
Mining microarray data at NCBI's Gene Expression Omnibus (GEO)\*.  
The use of PubMed/Medline in psychiatry. 3: Searching PubMed.  
Agency warns about dosing error for amphotericin after patients with cancer die.  
Electrophoretic recording of electronically stored radiographs.  
Multiparametric curve fitting-XIII Reliability of formation constants determined by analysis of potentiometric titration data.  
Issues in the measurement of social determinants of health.  
The neural crest.  
Transcatheter Amplatzer vascular plug-embolization of a giant postnephrectomy arteriovenous fistula combined with an aneurysm.  
Improved blood velocity measurements with a hybrid image filtering and iterative Radon transform algorithm.  
Reviewer acknowledgement.  
SAR11 Bacteria: The Most Abundant Plankton in the Oceans.

---

```
kable(zeroAuthors)
```

---

title

[150th Anniversary Celebration of the Royal Academy of Medicine of Belgium. Part 1. Bruxelles, 26-28 September 1991].  
Bibliography of toxinology.  
32nd Annual meeting of the Swiss Society of Nephrology. Lausanne, 14-15 December 2000. Abstracts.  
Information from your family doctor. Exercise for the elderly.  
Summaries for patients. Screening for osteoporosis: recommendations from the U.S. Preventive Services Task Force.  
Cancer death rates have fallen faster since 2002.  
BRUTALITIES of Nazi physicians.  
NA  
NA  
NA  
NA  
NA  
NA  
NA  
In results, there are 78 articles with 1 to 3 authors, 92 with "0" to 3 authors.

---

## Question 5

Find the average number of authors for articles. Display a single number.

```
mean(publication$authors)
```

```
## [1] 9.819876
```

The average number of authors for articles is almost 10.

## Question 6

What is the average time period (in days) between date completed and date revised. Display the time elapsed in days. Only consider cases where the difference is a positive number.

```
# Select articles that has positive difference between date completed and date revised
timePeriod <- publication %>%
  filter(dateComp < dateRevi) %>%
  select(issn,title, dateComp, dateRevi)

# Initialize the parameters
timeRow <- nrow(timePeriod)
timeInterval <- NA * timeRow
timePeriod["timeElapsed"] <- NA

# Loop through all rows
for (i in 1:timeRow) {
  timeInterval[i] <- interval(timePeriod$dateComp[i], timePeriod$dateRevi[i])
  timePeriod$timeElapsed[i] <- as.duration(timeInterval[i]) / ddays(1)
}

# display the table
kable(timePeriod)
```

issn	title
0095-3814	Hospital debt management and cost reimbursement.
0377-8231	[150th Anniversary Celebration of the Royal Academy of Medicine of Belgium. Part 1. Bruxelles, 26-28 September 1993].
0022-2623	3-Hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors. 7. Modification of the hexahydronaphthalene moiety.
0021-8820	Semisynthetic beta-lactam antibiotics. III. Effect on antibacterial activity and drug-susceptibility of chlorine- and amide-substituted derivatives.
0014-2956	Purification and characterisation of a water-soluble ferroxidase from <i>Bacillus subtilis</i> .
0017-0011	[Effect of fetal and neonatal growth on the occurrence of some diseases in adults].
0026-895X	Mechanism of extracellular ATP-induced proliferation of vascular smooth muscle cells.
1059-2725	Conservative management of mechanical neck disorders. A systematic overview and meta-analysis.
0009-7322	Assessment of myocardial viability in patients with chronic coronary artery disease. Rest-4-hour-24-hour 201Tl myocardial perfusion scintigraphy.
0014-2980	Hypermutation, diversity and dissemination of human intestinal lamina propria plasma cells.
0140-6736	A woman with nodules in her lungs.
0007-0912	Myocardial ischaemia after coronary artery bypass grafting: early vs late extubation.
0002-838X	Lumbar spine stenosis: a common cause of back and leg pain.
1435-2443	Outcome of patients with sepsis and septic shock after ICU treatment.
0022-1899	Retraction: A rabbit model for human cytomegalovirus-induced chorioretinal disease (J Infect Dis 1993;168:336-340).
0047-1828	Hepatitis C virus infection and heart diseases: a multicenter study in Japan.
0268-1315	Cardiac side-effects of two selective serotonin reuptake inhibitors in middle-aged and elderly depressed patients.
0300-2896	[Tobacco control in children, adolescents and young people: knowledge, prevention and action].
0893-133X	Antipsychotic potential of CCK-based treatments: an assessment using the prepulse inhibition model of psychosis.
0041-0101	Bibliography of toxinology.
0832-610X	Complete airway obstruction.
0733-8627	Evaluation of the patient with extremity trauma: an evidence based approach.
0168-8510	Health technology assessment: decentralized and fragmented in the US compared to other countries.
0027-8424	Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic bacteria.
0031-7144	Antimicrobial activity of some Nepalese medicinal plants.

issn	title
0031-7144	Different kinetics of hydroquinone depletion in various medicinal plant tissue cultures producing arbutin.
0031-9023	Looking for the Forrest.
0031-9023	Effects of side lying on lung function in older individuals.
0869-8139	[Ethanol modifies the ion selectivity of sodium channels in the rat sensory neurons].
0950-382X	Transcription regulation of the nir gene cluster encoding nitrite reductase of <i>Paracoccus denitrificans</i> involves
1432-2218	Inflammatory fibroid polyp of the duodenum.
0022-2623	Phosphorylated morpholine acetal human neurokinin-1 receptor antagonists as water-soluble prodrugs.
1098-1004	Erratum: analysis of DNA elements that modulate myosin VIIa expression in humans.
0899-1987	Altered expression of BRCA1, BRCA2, and a newly identified BRCA2 exon 12 deletion variant in malignant h
0108-2701	Multicentre hydrogen bonds in a 2:1 arylsulfonylimidazolone hydrochloride salt.
1469-493X	Parent-training programmes for improving maternal psychosocial health.
0022-1767	Histamine induces exocytosis and IL-6 production from human lung macrophages through interaction with H1
1469-493X	Elective versus selective caesarean section for delivery of the small baby.
0195-668X	Indications for implantable cardioverter defibrillator (ICD) therapy. Study Group on Guidelines on ICDs of th
1469-493X	Parent-training programmes for improving maternal psychosocial health.
0284-186X	Assessment of quality of life during chemotherapy.
0303-6979	Utilisation of locally delivered doxycycline in non-surgical treatment of chronic periodontitis. A comparative m
0028-0836	Initial sequencing and analysis of the human genome.
0905-7161	Early endosseous integration enhanced by dual acid etching of titanium: a torque removal study in the rabbit.
0021-9258	Evidence that ligand and metal ion binding to integrin alpha 4beta 1 are regulated through a coupled equilibri
0905-7161	Histology of human alveolar bone regeneration with a porous tricalcium phosphate. A report of two cases.
0022-2623	Novel azo derivatives as prodrugs of 5-aminosalicylic acid and amino derivatives with potent platelet activating
0250-5525	32nd Annual meeting of the Swiss Society of Nephrology. Lausanne, 14-15 December 2000. Abstracts.
1098-1004	Detection of six novel FBN1 mutations in British patients affected by Marfan syndrome.
0033-183X	Cytoplasmic calcium levels in protoplasts from the cap and elongation zone of maize roots.
0735-2689	Calcium messenger system in plants.
0198-8859	The origin of Palestinians and their genetic relatedness with other Mediterranean populations.
0036-8075	Changes in seismic anisotropy after volcanic eruptions: evidence from Mount Ruapehu.
0031-9007	Maximal height scaling of kinetically growing surfaces.
0084-6597	Understanding oblique impacts from experiments, observations, and modeling.
0198-8859	Letter from the ASHI president and council.
0032-4728	On the origins of the United States Government's international population policy.
0511-0726	Demographic transition in the Netherlands: a statistical analysis of regional differences in the level and develop
0024-2160	Thomas Dover's "Ancient physician's legacy".
0098-7484	msJAMA: Breast reconstruction: one woman's choice.
0028-4793	Recognition and management of anthrax—an update.
0031-7012	Progress in the search for ideal drugs.
1098-1004	Erratum: Detection of six novel FBN1 mutations in British patients affected by Marfan syndrome.
1098-1004	Eight novel germline MLH1 and MSH2 mutations in hereditary non-polyposis colorectal cancer families from S
0003-9950	Republication of color figures.
0023-7205	[Unequal distribution of health resources. Research is not supported with consideration to the global significan
0002-838X	Information from your family doctor. Exercise for the elderly.
1052-1577	By the way doctor... I'm 72 and was recently told that I am losing a bit of my sight because of macular dege
1069-9422	Consumer acceptance of vegetarian sweet potato products intended for space missions.
0028-2200	[Microbiological diagnostics in periodontal treatment].
0012-1606	The major subacrosomal occupant of bull spermatozoa is a novel histone H2B variant associated with the form
1070-910X	By the way, doctor. A few months ago, I was surprised to read that a new study has raised a question about t
1432-2218	Disruptive visions: surgeon responsibility during the era of change.
0098-7484	The 2001 Bethesda System: terminology for reporting results of cervical cytology.
0273-1177	Planetary protection issues for Mars sample acquisition flight projects.
0021-9258	Interaction between FtsZ and FtsW of <i>Mycobacterium tuberculosis</i> .
0362-4331	Weighing medical ethics for many years to come: a conversation with Harold Shapiro. Interview by Howard M



issn	title
1145-0762	Physicians' attitudes towards medical ethics issues in Turkey.
0021-9258	Amisyn, a novel syntaxin-binding protein that may regulate SNARE complex assembly.
0018-9251	Intelligent control of a planning system for astronaut training.
1226-0282	Recent changes in the population control policy and its future directions in Korea.
1095-9203	Mycobacterium leprae and demyelination.
1054-6863	Public policy and the sale of human organs.
1225-505X	Development of neurophysiology in the early twentieth century: Charles Scott Sherrington and The Integrative
0269-8897	Paracelsus, Paracelsianism, and the secularization of the worldview.
1539-3704	Summaries for patients. Screening for osteoporosis: recommendations from the U.S. Preventive Services Task I
1539-3704	Screening for osteoporosis in postmenopausal women: recommendations and rationale.
0251-7329	Gender and globalization. A century in retrospect.
8750-7587	Human unilateral lower limb suspension as a model for spaceflight effects on skeletal muscle.
0559-7765	[Mechanism of bone mineral loss in microgravity].
0025-4282	The new Uniform Health Care Decisions Act: paving a health care decisions superhighway?
0025-4282	The right to refuse life-sustaining medical treatment: national trend and recent changes in Maryland law.
0031-9007	Measurement of B -> K*gamma branching fractions and charge asymmetries.
1879-0712	S-15176 inhibits mitochondrial permeability transition via a mechanism independent of its antioxidant property
1879-0712	Risperidone reduces limited access alcohol drinking in alcohol-preferring rats.
1879-0712	Behavioral effects of rimcazole analogues alone and in combination with cocaine.
0014-2565	[Physiopathology of phosphorus and calcium changes and of bone lesions in glomerular nephropathies].
0016-5662	[JEJUNAL BIOPSY AND LIPEMIA TOLERANCE TEST IN CIRRHOTIC PATIENTS].
0002-9378	SEROTONIN CONTENT OF HUMAN PLACENTA AND FETUS DURING PREGNANCY.
1087-2108	Epidermal nevus.
1226-0479	[The significance of urine sodium measurement after furosemide administration in diuretics-unresponsive patien
1362-4962	Local homology recognition and distance measures in linear time using compressed amino acid alphabets.
1539-6150	Hampering a heartbreaker. Antibiotic might stem injury from heart attack.
1362-4962	Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter.
0028-0836	Reduction by molecular hydrogen of acetoacetate to butyrate by butyric acid bacteria.
0364-2348	Spontaneous resolution of solitary osteochondroma in the young adult.
0968-0896	Aminyl and iminyl radicals from arylhydrazones in the photo-induced DNA cleavage.
0896-4289	Warm partner contact is related to lower cardiovascular reactivity.
1080-6032	Acute coronary ischemia following centipede envenomation: case report and review of the literature.
0913-8668	Hypoxic ventilatory response in cats lightly anesthetized with ketamine: effects of halothane and sevoflurane in
0027-8424	Tubular precipitation and redox gradients on a bubbling template.
1551-4005	Altered epigenetic patterning leading to replicative senescence and reduced longevity. A role of a novel SNF2 f
1551-4005	TrkAIII. A novel hypoxia-regulated alternative TrkA splice variant of potential physiological and pathological
1551-4005	Replication timing of human chromosome 6.
1000-8713	[The determination of olaquinox in feeds by HPLC].
0248-4900	Books for free? How can this be? - A PubMed resource you may be overlooking.
1539-3704	The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic
1942-597X	MeSH Speller + askMEDLINE: auto-completes MeSH terms then searches MEDLINE/PubMed via free-text, r
1064-3745	Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*.
1432-2218	Crura ultrastructural alterations in patients with hiatal hernia: a pilot study.
1359-6535	Declining prevalence of HIV-1 drug resistance in treatment-failing patients: a clinical cohort study.
1756-1833	Agency warns about dosing error for amphotericin after patients with cancer die.
1042-8194	Advanced age and high initial WBC influence the outcome of inv(3) (q21q26)/t(3;3) (q21;q26) positive AML.
1091-0220	Cancer death rates have fallen faster since 2002.
0002-9955	BRUTALITIES of Nazi physicians.
1756-1833	Health related quality of life after combined hormone replacement therapy: randomised controlled trial.
1477-9129	The neural crest.
1612-3174	Transcatheter Amplatzer vascular plug-embolization of a giant postnephrectomy arteriovenous fistula combined
1662-4548	Improved blood velocity measurements with a hybrid image filtering and iterative Radon transform algorithm.

issn	title
1460-2075	Eps8 controls dendritic spine density and synaptic plasticity through its actin-capping activity.
2157-3999	Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions.
1749-7922	Reviewer acknowledgement.
1529-2401	A new pathway mediating social effects on the endocrine system: female presence acting via norepinephrine re

```
# average time period
mean(timePeriod$timeElapsed)
```

```
## [1] 3796.286
```

The average time period between date completed and date revised is 3796 days.

## Question 7

Which articles published in PubMed were not written in English? Only consider the first language of publication.

```
nonEng <- journal %>%
  filter(lang != "eng") %>%
  left_join(publication, by = c("issn", "pubYear")) %>%
  select(jourTitle, title, lang)
kable(nonEng)
```

jourTitle	title
Bulletin et memoires de l'Academie royale de medecine de Belgique	[150th Anniversary Celebration of the Royal Academy of Medicine]
Ginekologia polska	[Effect of fetal and neonatal growth on the occurrence of preeclampsia]
Archivos de bronconeumologia	[Tobacco control in children, adolescents and young adults]
Rossiiskii fiziologicheskii zhurnal imeni I.M. Sechenova	[Ethanol modifies the ion selectivity of sodium channels]
Lakartidningen	[Unequal distribution of health resources. Research results from the Swedish Health Care Survey]
Nederlands tijdschrift voor tandheelkunde	[Microbiological diagnostics in periodontal treatment]
Sheng li ke xue jin zhan [Progress in physiology]	[Mechanism of bone mineral loss in microgravity].
Revista clinica espanola	[Physiopathology of phosphorus and calcium changes in chronic kidney disease]
Gazzetta internazionale di medicina e chirurgia	[JEJUNAL BIOPSY AND LIPEMIA TOLERANCE IN THE RAT]
Taehan Kan Hakhoe chi = The Korean journal of hepatology	[The significance of urine sodium measurement after liver transplantation]
Se pu = Chinese journal of chromatography	[The determination of olaquinox in feeds by HPLC]
Guang pu xue yu guang pu fen xi = Guang pu	[Study on cleavage mechanisms of dimer (t-Bu) <sub>2</sub> NCO]
Tijdschrift voor gerontologie en geriatrie	NA
Tijdschrift voor gerontologie en geriatrie	NA
Tijdschrift voor gerontologie en geriatrie	NA
Tijdschrift voor gerontologie en geriatrie	NA
Tijdschrift voor gerontologie en geriatrie	NA
Tijdschrift voor gerontologie en geriatrie	NA
Tijdschrift voor gerontologie en geriatrie	NA
The table above listed out the articles published in PubMed were not	written in English.

## Question 8

Using the XML data (not the tibbles created above), find the articles containing any of the words “drug resistance” or “virus” in any capitalization in the title. Note that drug resistance could be spelled as “drug resistance” or “drug-resistance” or “drug resistant” or “drug resistant” – use regular expressions to deal with the variations.

```
med01 <- readLines("http://da5020.weebly.com/uploads/8/6/5/9/8659576/pubmedsample.jun18.xml")

drug <- regexpr("<ArticleTitle>(.*)[D|d]rug(.*)[R|r]esist(.*?)</ArticleTitle>", med01)#regular expression
d <- regmatches(med01, drug)#saves matched terms as a vector.
(drugTitles <- gsub("<ArticleTitle>|</ArticleTitle>", "", d))#removes the "<ArticleTitle>" tags so only

## [1] "Declining prevalence of HIV-1 drug resistance in treatment-failing patients: a clinical cohort s

virus <- regexpr("<ArticleTitle>(.*)[V|v]irus(.*?)</ArticleTitle>", med01)#regular expression that finds
v <- regmatches(med01, virus)#saves matched terms as a vector.
(virusTitles <- gsub("<ArticleTitle>|</ArticleTitle>", "", v))#removes the "<ArticleTitle>" tags so only

## [1] "Retraction: A rabbit model for human cytomegalovirus--induced chorioretinal disease (J Infect Dis
## [2] "Hepatitis C virus infection and heart diseases: a multicenter study in Japan."
```

The listed 3 articles containing any of the words “drug resistance” or “virus”.