

## TextLink\_LabNotes01

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

### Navigation

Ctrl+Alt + i to insert new r code chunk

Your current working directory

```
getwd()

## [1] "/home/cinkova"
```

Data are in edu/r/textlink/src\_data, File "PDT\_30.csv". Create path to data set to reuse, save into variable src\_path

```
src_path <- file.path("edu", "r", "textlink", "src_data")
```

Create string of path to the "PDT\_30.csv" file, use function "paste" with help (type ?paste in Console and hit Enter). Tweak the string and save it to path\_pdt30.

```
path_pdt30 <- paste(src_path, "PDT_30.csv", sep = "/")
```

### File Import

Hazard format guess: read\_csv, read\_delim, read\_csv2, scan (use parameters what = "character" and n = 5 with scan), or readLines (use parameter n = 5 with readLines to read just 5 rows), Check result in GUI. For printing to screen restrict the number of lines in read\_csv/csv2/delim to 5 by nrow = 5.

```
read.csv(file = path_pdt30, nrow = 5)

##   document_id  genre number_of_sentences  sentence_id
## 1 cmpr9410_001 comment                   33 t-cmpr9410-001-p10s2
```

```
## 2 cmpr9410_001 comment          33 t-cmpr9410-001-p11s2
## 3 cmpr9410_001 comment          33 t-cmpr9410-001-p15s1
## 4 cmpr9410_001 comment          33 t-cmpr9410-001-p17s1
## 5 cmpr9410_001 comment          33 t-cmpr9410-001-p17s1
## discourse_type discourse_class
## 1 conc CONTRAST
## 2 opp CONTRAST
## 3 spec EXPANSION
## 4 restr CONTRAST
## 5 conj EXPANSION
```

Read the entire data file and store it in variable pdt30.

```
pdt30 <- read.csv(file = path_pdt30)
```

## Basic Exploration

Look at the first/last/random n lines with functions head, tail, dplyr::sample\_n. Install library dplyr to use the dplyr::sample\_n function and set seed to 123 make your random selection reproducible.

```
head(pdt30)
```

```
## document_id genre number_of_sentences sentence_id
## 1 cmpr9410_001 comment          33 t-cmpr9410-001-p10s2
## 2 cmpr9410_001 comment          33 t-cmpr9410-001-p11s2
## 3 cmpr9410_001 comment          33 t-cmpr9410-001-p15s1
## 4 cmpr9410_001 comment          33 t-cmpr9410-001-p17s1
## 5 cmpr9410_001 comment          33 t-cmpr9410-001-p17s1
## 6 cmpr9410_001 comment          33 t-cmpr9410-001-p24s2
## discourse_type discourse_class
## 1 conc CONTRAST
## 2 opp CONTRAST
## 3 spec EXPANSION
## 4 restr CONTRAST
## 5 conj EXPANSION
## 6 conc CONTRAST
```

```
tail(pdt30)
```

```
## document_id genre number_of_sentences sentence_id
## 20551 mf930713_161 news          24 t-mf930713-161-p2s7
## 20552 mf930713_162 news          15 t-mf930713-162-p2s5
## 20553 mf930713_162 news          15 t-mf930713-162-p2s9
## 20554 mf930713_163 collection    19 t-mf930713-163-p1s1C
## 20555 mf930713_163 collection    19 t-mf930713-163-p1s5
## 20556 mf930713_163 collection    19 t-mf930713-163-p2s3
## discourse_type discourse_class
## 20551 reason CONTINGENCY
## 20552 conj EXPANSION
```

```
## 20553      spec      EXPANSION
## 20554      conc      CONTRAST
## 20555      preced    TEMPORAL
## 20556      opp       CONTRAST
```

```
library(dplyr)
```

```
set.seed(1234)
```

```
dplyr::sample_n(pdt30, 10)
```

```
##      document_id      genre number_of_sentences      sentence_id
## 2338  ln94200_155      sport                45  t-ln94200-155-p4s5
## 12792 ln95045_017      news                 12  t-ln95045-017-p3s1
## 12524  ln94211_92      essay                87  t-ln94211-92-p13s6
## 12813  ln95045_023      overview            14  t-ln95045-023-p3s1
## 17694 lnd94103_122      sport                17  t-lnd94103-122-p1s12
## 13160  ln95045_077      news                 27  t-ln95045-077-p6s1
## 196    cmpr9410_019    topic_interv        91  t-cmpr9410-019-p21s1
## 4779   ln94203_85      essay               133  t-ln94203-85-p6s1
## 13687  ln95046_076      essay                29  t-ln95046-076-p2s9
## 10567  ln94209_49    topic_interv        41  t-ln94209-49-p3s2
##      discourse_type discourse_class
## 2338      conj      EXPANSION
## 12792      conj      EXPANSION
## 12524      purp     CONTINGENCY
## 12813      conj      EXPANSION
## 17694      opp      CONTRAST
## 13160      conj      EXPANSION
## 196        opp      CONTRAST
## 4779      conj      EXPANSION
## 13687      f_opp    CONTRAST
## 10567      reason   CONTINGENCY
```

Use `str` to learn more about `pdt30`

```
str(pdt30)
```

```
## 'data.frame':  20556 obs. of  6 variables:
## $ document_id      : Factor w/ 2580 levels "cmpr9410_001",...: 1 1 1 1 1
1 1 1 1 1 ...
## $ genre            : Factor w/ 19 levels "advice","caption",...: 4 4 4 4
4 4 4 4 4 ...
## $ number_of_sentences: int  33 33 33 33 33 33 33 33 33 33 ...
## $ sentence_id      : Factor w/ 16112 levels "t-cmpr9410-001-p10s2",...:
1 2 3 4 4 5 6 6 7 8 ...
## $ discourse_type    : Factor w/ 23 levels "conc","cond",...: 1 16 22 21 4
1 21 4 2 16 ...
## $ discourse_class   : Factor w/ 4 levels "CONTINGENCY",...: 2 2 3 2 3 2 2
3 1 2 ...
```

Use `dplyr::glimpse` to learn more about `pdt30`

```
dplyr::glimpse(pdt30)

## Observations: 20,556
## Variables: 6
## $ document_id      <fctr> cmpr9410_001, cmpr9410_001, cmpr9410_001,...
## $ genre            <fctr> comment, comment, comment, comment, comme...
## $ number_of_sentences <int> 33, 33, 33, 33, 33, 33, 33, 33, 33, 33...
## $ sentence_id      <fctr> t-cmpr9410-001-p10s2, t-cmpr9410-001-p11s...
## $ discourse_type    <fctr> conc, opp, spec, restr, conj, conc, restr...
## $ discourse_class   <fctr> CONTRAST, CONTRAST, EXPANSION, CONTRAST, ...
```

Use summary to learn more about pdt30. Mind empty values!

```
summary(pdt30)

##      document_id      genre      number_of_sentences
## ln94207_73 : 123 news      :4510 Min. : 1.0
## ln94207_76 : 118 essay      :3757 1st Qu.: 16.0
## ln94210_95 : 102 sport      :2525 Median : 26.0
## ln94211_92 : 87 description :2305 Mean : 37.3
## ln94200_127: 80 comment     :1731 3rd Qu.: 42.0
## ln94207_79 : 77 topic_interv:1247 Max. : 231.0
## (Other) :19969 (Other) :4481
##      sentence_id      discourse_type      discourse_class
## t-ln94203-125-p8s11 : 6 conj :7498 CONTINGENCY:4701
## t-ln94207-73-p13s7 : 6 opp :3196 CONTRAST :5938
## t-ln94207-90-p4s9 : 6 reason :2632 EXPANSION :8851
## t-ln95048-061-p7s3 : 6 cond :1369 TEMPORAL :1066
## t-cmpr9415-033-p11s4: 5 conc : 880
## t-ln94203-99-p8s1 : 5 preced : 840
## (Other) :20522 (Other):4141
```

Examine some columns with categorical variables: do their categories contain what you have expected? use levels and subset the pdt30 data frame with both column names and column indices

```
levels(pdt30$genre)

## [1] "advice"      "caption"      "collection"    "comment"
## [5] "description" "essay"        "invitation"    "letter"
## [9] "news"        "other"        "overview"      "person_interv"
## [13] "plot"        "program"      "review"        "sport"
## [17] "survey"      "topic_interv" "weather"
```

```
head(pdt30[1,1])
```

```
## [1] cmpr9410_001
## 2580 Levels: cmpr9410_001 cmpr9410_002 cmpr9410_003 ... mf930713_163
```

```
head(pdt30[,2])
```

```
## [1] comment comment comment comment comment comment
## 19 Levels: advice caption collection comment description ... weather

head(pdt30[1, c(2,5)])

##      genre discourse_type
## 1 comment                conc
```

Get column names of the data frame (useful with very wide data sets, not our case)

```
colnames(pdt30)

## [1] "document_id"          "genre"                "number_of_sentences"
## [4] "sentence_id"         "discourse_type"      "discourse_class"
```

## Subsetting

Get rows which meet conditions you set. Use `dplyr::filter`.

Find rows with genre comment and number of sentences less than 40. Display only the head of the result

```
head(dplyr::filter(pdt30, genre == "comment" & number_of_sentences < 40))

##   document_id  genre number_of_sentences      sentence_id
## 1 cmpr9410_001 comment                33 t-cmpr9410-001-p10s2
## 2 cmpr9410_001 comment                33 t-cmpr9410-001-p11s2
## 3 cmpr9410_001 comment                33 t-cmpr9410-001-p15s1
## 4 cmpr9410_001 comment                33 t-cmpr9410-001-p17s1
## 5 cmpr9410_001 comment                33 t-cmpr9410-001-p17s1
## 6 cmpr9410_001 comment                33 t-cmpr9410-001-p24s2
##   discourse_type discourse_class
## 1              conc          CONTRAST
## 2              opp          CONTRAST
## 3              spec          EXPANSION
## 4             restr          CONTRAST
## 5              conj          EXPANSION
## 6              conc          CONTRAST
```

Find rows with genre weather or plot, respectively, use `|` and `%in%`, respectively, in your logical expressions

```
dplyr::filter(pdt30, genre == "weather" | genre == "plot")

##   document_id  genre number_of_sentences      sentence_id
## 1 ln94200_10   plot                10 t-ln94200-10-p3s2
## 2 ln94200_11   plot                 6 t-ln94200-11-p2s2
## 3 ln94200_14   plot                 5 t-ln94200-14-p2s1
## 4 ln94200_15   plot                 6 t-ln94200-15-p2s1
## 5 ln94200_16   plot                 8 t-ln94200-16-p3s1
## 6 ln94200_16   plot                 8 t-ln94200-16-p3s2
## 7 ln94200_16   plot                 8 t-ln94200-16-p3s2
```

## 8	ln94200_16	plot	8	t-ln94200-16-p3s2
## 9	ln94200_16	plot	8	t-ln94200-16-p3s2
## 10	ln94200_16	plot	8	t-ln94200-16-p4s2
## 11	ln94200_17	plot	6	t-ln94200-17-p2s2
## 12	ln94200_2	plot	5	t-ln94200-2-p2s3
## 13	ln94207_113	plot	15	t-ln94207-113-p2s2
## 14	ln94207_113	plot	15	t-ln94207-113-p3s1
## 15	ln94207_113	plot	15	t-ln94207-113-p3s3
## 16	ln95047_043	plot	8	t-ln95047-043-p2s1
## 17	ln95047_043	plot	8	t-ln95047-043-p2s2
## 18	ln95047_043	plot	8	t-ln95047-043-p2s3A
## 19	ln95047_043	plot	8	t-ln95047-043-p2s3A
## 20	ln95047_043	plot	8	t-ln95047-043-p2s4
## 21	ln95047_043	plot	8	t-ln95047-043-p3s1
## 22	ln95047_053	plot	4	t-ln95047-053-p2s2
## 23	lnd94103_129	weather	9	t-lnd94103-129-p1s10
## 24	lnd94103_129	weather	9	t-lnd94103-129-p1s9
## 25	lnd94103_153	weather	8	t-lnd94103-153-p1s6A
## 26	lnd94103_153	weather	8	t-lnd94103-153-p1s7
## 27	mf920922_043	weather	14	t-mf920922-043-p3s1
## 28	mf920922_043	weather	14	t-mf920922-043-p3s1
## 29	mf920925_027	weather	11	t-mf920925-027-p3s1
## 30	mf920925_027	weather	11	t-mf920925-027-p3s1
## 31	mf930709_066	weather	48	t-mf930709-066-p4s2
## 32	mf930709_066	weather	48	t-mf930709-066-p4s2
## 33	mf930709_066	weather	48	t-mf930709-066-p6s2B
## 34	mf930709_066	weather	48	t-mf930709-066-p8s1
## 35	mf930713_099	weather	15	t-mf930713-099-p4s2
## 36	mf930713_099	weather	15	t-mf930713-099-p4s2
##	discourse_type	discourse_class		
## 1	opp	CONTRAST		
## 2	conj	EXPANSION		
## 3	conj	EXPANSION		
## 4	conc	CONTRAST		
## 5	conj	EXPANSION		
## 6	reason	CONTINGENCY		
## 7	preced	TEMPORAL		
## 8	opp	CONTRAST		
## 9	reason	CONTINGENCY		
## 10	corr	CONTRAST		
## 11	conj	EXPANSION		
## 12	conj	EXPANSION		
## 13	cond	CONTINGENCY		
## 14	conj	EXPANSION		
## 15	conc	CONTRAST		
## 16	conj	EXPANSION		
## 17	purp	CONTINGENCY		
## 18	conj	EXPANSION		
## 19	conj	EXPANSION		
## 20	cond	CONTINGENCY		

```

## 21      conj      EXPANSION
## 22    preced      TEMPORAL
## 23      conj      EXPANSION
## 24      conj      EXPANSION
## 25      conj      EXPANSION
## 26      conj      EXPANSION
## 27      conj      EXPANSION
## 28      conj      EXPANSION
## 29      conj      EXPANSION
## 30      conj      EXPANSION
## 31      conj      EXPANSION
## 32      conj      EXPANSION
## 33      spec      EXPANSION
## 34      spec      EXPANSION
## 35      conj      EXPANSION
## 36      conj      EXPANSION

```

```
dplyr::filter(pdt30, genre %in% c("weather", "plot"))
```

```

##      document_id  genre number_of_sentences  sentence_id
## 1    ln94200_10    plot                10    t-ln94200-10-p3s2
## 2    ln94200_11    plot                 6    t-ln94200-11-p2s2
## 3    ln94200_14    plot                 5    t-ln94200-14-p2s1
## 4    ln94200_15    plot                 6    t-ln94200-15-p2s1
## 5    ln94200_16    plot                 8    t-ln94200-16-p3s1
## 6    ln94200_16    plot                 8    t-ln94200-16-p3s2
## 7    ln94200_16    plot                 8    t-ln94200-16-p3s2
## 8    ln94200_16    plot                 8    t-ln94200-16-p3s2
## 9    ln94200_16    plot                 8    t-ln94200-16-p3s2
## 10   ln94200_16    plot                 8    t-ln94200-16-p4s2
## 11   ln94200_17    plot                 6    t-ln94200-17-p2s2
## 12   ln94200_2     plot                 5    t-ln94200-2-p2s3
## 13   ln94207_113   plot                15    t-ln94207-113-p2s2
## 14   ln94207_113   plot                15    t-ln94207-113-p3s1
## 15   ln94207_113   plot                15    t-ln94207-113-p3s3
## 16   ln95047_043   plot                 8    t-ln95047-043-p2s1
## 17   ln95047_043   plot                 8    t-ln95047-043-p2s2
## 18   ln95047_043   plot                 8    t-ln95047-043-p2s3A
## 19   ln95047_043   plot                 8    t-ln95047-043-p2s3A
## 20   ln95047_043   plot                 8    t-ln95047-043-p2s4
## 21   ln95047_043   plot                 8    t-ln95047-043-p3s1
## 22   ln95047_053   plot                 4    t-ln95047-053-p2s2
## 23   lnd94103_129   weather                9    t-lnd94103-129-p1s10
## 24   lnd94103_129   weather                9    t-lnd94103-129-p1s9
## 25   lnd94103_153   weather                8    t-lnd94103-153-p1s6A
## 26   lnd94103_153   weather                8    t-lnd94103-153-p1s7
## 27   mf920922_043   weather                14    t-mf920922-043-p3s1
## 28   mf920922_043   weather                14    t-mf920922-043-p3s1
## 29   mf920925_027   weather                11    t-mf920925-027-p3s1
## 30   mf920925_027   weather                11    t-mf920925-027-p3s1

```

```

## 31 mf930709_066 weather 48 t-mf930709-066-p4s2
## 32 mf930709_066 weather 48 t-mf930709-066-p4s2
## 33 mf930709_066 weather 48 t-mf930709-066-p6s2B
## 34 mf930709_066 weather 48 t-mf930709-066-p8s1
## 35 mf930713_099 weather 15 t-mf930713-099-p4s2
## 36 mf930713_099 weather 15 t-mf930713-099-p4s2
## discourse_type discourse_class
## 1 opp CONTRAST
## 2 conj EXPANSION
## 3 conj EXPANSION
## 4 conc CONTRAST
## 5 conj EXPANSION
## 6 reason CONTINGENCY
## 7 preced TEMPORAL
## 8 opp CONTRAST
## 9 reason CONTINGENCY
## 10 corr CONTRAST
## 11 conj EXPANSION
## 12 conj EXPANSION
## 13 cond CONTINGENCY
## 14 conj EXPANSION
## 15 conc CONTRAST
## 16 conj EXPANSION
## 17 purp CONTINGENCY
## 18 conj EXPANSION
## 19 conj EXPANSION
## 20 cond CONTINGENCY
## 21 conj EXPANSION
## 22 preced TEMPORAL
## 23 conj EXPANSION
## 24 conj EXPANSION
## 25 conj EXPANSION
## 26 conj EXPANSION
## 27 conj EXPANSION
## 28 conj EXPANSION
## 29 conj EXPANSION
## 30 conj EXPANSION
## 31 conj EXPANSION
## 32 conj EXPANSION
## 33 spec EXPANSION
## 34 spec EXPANSION
## 35 conj EXPANSION
## 36 conj EXPANSION

```

Remove column `discourse_type` from `pdt30`, use `dplyr::select`, store in variable `pdt_small`, view first 10 rows.

```

pdt_small <- dplyr::select(pdt30, -discourse_type)
head(pdt_small, 10)

```



```

##   document_id  genre number_of_sentences  sentence_id
## 1 cmpr9410_001 comment                33 t-cmpr9410-001-p10s2
## 2 cmpr9410_001 comment                33 t-cmpr9410-001-p11s2
## 3 cmpr9410_001 comment                33 t-cmpr9410-001-p15s1
## 4 cmpr9410_001 comment                33 t-cmpr9410-001-p17s1
## 5 cmpr9410_001 comment                33 t-cmpr9410-001-p17s1
## 6 cmpr9410_001 comment                33 t-cmpr9410-001-p24s2
## 7 cmpr9410_001 comment                33 t-cmpr9410-001-p4s2
## 8 cmpr9410_001 comment                33 t-cmpr9410-001-p4s2
## 9 cmpr9410_001 comment                33 t-cmpr9410-001-p4s3
##10 cmpr9410_001 comment                33 t-cmpr9410-001-p8s1
##   discourse_class
## 1          CONTRAST
## 2          CONTRAST
## 3      EXPANSION
## 4          CONTRAST
## 5      EXPANSION
## 6          CONTRAST
## 7          CONTRAST
## 8      EXPANSION
## 9  CONTINGENCY
##10          CONTRAST

```

Select columns `document_id` and `sentence_id` from `pdt_small`. Use `ends_with`. Save as `pdt_smaller`. See head.

```

pdt_smaller <- dplyr::select(pdt_small, ends_with("_id"))
head(pdt_smaller)

```

```

##   document_id  sentence_id
## 1 cmpr9410_001 t-cmpr9410-001-p10s2
## 2 cmpr9410_001 t-cmpr9410-001-p11s2
## 3 cmpr9410_001 t-cmpr9410-001-p15s1
## 4 cmpr9410_001 t-cmpr9410-001-p17s1
## 5 cmpr9410_001 t-cmpr9410-001-p17s1
## 6 cmpr9410_001 t-cmpr9410-001-p24s2

```

Remove columns `document_id` and `sentence_id` from `pdt_small`. Use `ends_with`. Save as `pdt_smallest`. See head.

```

pdt_smallest <- dplyr::select(pdt_small, -(ends_with("_id")))
head(pdt_smallest)

```

```

##   genre number_of_sentences  discourse_class
## 1 comment                33          CONTRAST
## 2 comment                33          CONTRAST
## 3 comment                33      EXPANSION
## 4 comment                33          CONTRAST
## 5 comment                33      EXPANSION
## 6 comment                33          CONTRAST

```

Remove pdt\_small from your workspace.

```
rm(pdt_small)
```

Remove the other ones using a regular expression pattern in ls.

```
rm(list = ls(pattern = "small"))
```

Remember the outlier document in collection that was very long and only contained one connective. Let's get rid of it. Filter the data set and remove the row in which genre equals collection and number\_of\_sentences is greater than 150. Watch the Environment panel. pdt30 will change "Value" from 20556 to 20555 obs. of 6 variables.

```
pdt30 <- dplyr::filter(pdt30, !(genre == "collection" & number_of_sentences > 150))
```

## Factor Levels & Editing

Now, again in pdt30, merge the genre categories "weather", "plot", "caption", and "program". How-to: Check the unique categories, which are levels of the factor pdt30\$genre. Replace these levels with level "other". Check levels now.

```
levels(pdt30$genre)[levels(pdt30$genre) %in% c("weather", "plot", "caption", "program")] <- "other"  
levels(pdt30$genre)
```

```
## [1] "advice"      "other"      [1] "collection"  "comment"  
## [5] "description" "essay"     [5] "invitation"  "letter"  
## [9] "news"       "overview"  [9] "person_interv" "review"  
## [13] "sport"     "survey"   [13] "topic_interv"
```

## Stats

Cross-tabulate genres vs. discourse\_type, by discourse\_class, use table

```
table(pdt30$genre, pdt30$discourse_type, pdt30$discourse_class)
```

```
## , , = CONTINGENCY  
##  
##  
##          conc cond confr conj conjalt corr disjalt equiv exempl  
## advice          0 102    0  0         0  0         0  0         0  
## other            0  31    0  0         0  0         0  0         0  
## collection      0  24    0  0         0  0         0  0         0  
## comment         0 149    0  0         0  0         0  0         0  
## description     0 156    0  0         0  0         0  0         0  
## essay           0 278    0  0         0  0         0  0         0  
## invitation      0  7     0  0         0  0         0  0         0  
## letter          0  31    0  0         0  0         0  0         0
```

```

## news          0 231    0  0    0  0    0  0    0
## overview      0  10    0  0    0  0    0  0    0
## person_interv 0  61    0  0    0  0    0  0    0
## review        0  51    0  0    0  0    0  0    0
## sport         0 112    0  0    0  0    0  0    0
## survey        0  14    0  0    0  0    0  0    0
## topic_interv  0 112    0  0    0  0    0  0    0

```

```

##
##          explicat f_cond f_opp f_reason gener grad opp other
## advice          6     0     0     2     0     0     0     0
## other           2     0     0     2     0     0     0     0
## collection      9     0     0     0     0     0     0     0
## comment        34     1     0     2     0     0     0     0
## description    17     1     0     2     0     0     0     0
## essay          54     7     0    15     0     0     0     0
## invitation      6     0     0     1     0     0     0     0
## letter          5     0     0     0     0     0     0     0
## news           36     3     0     8     0     0     0     0
## overview        0     0     0     1     0     0     0     0
## person_interv   4     2     0     3     0     0     0     0
## review         23     0     0     1     0     0     0     0
## sport          27     1     0     2     0     0     0     0
## survey          0     0     0     0     0     0     0     0
## topic_interv    7     1     0     1     0     0     0     0

```

```

##
##          preced purp reason restr spec synchr
## advice          0  15  117     0     0     0
## other           0  17   69     0     0     0
## collection      0  17   65     0     0     0
## comment         0  39  258     0     0     0
## description     0  39  305     0     0     0
## essay           0  77  482     0     0     0
## invitation      0   1   16     0     0     0
## letter          0   6   48     0     0     0
## news            0  97  608     0     0     0
## overview        0   0    7     0     0     0
## person_interv   0  22   81     0     0     0
## review          0  24  110     0     0     0
## sport           0  33  289     0     0     0
## survey          0   1   18     0     0     0
## topic_interv    0  26  159     0     0     0

```

```

##
## , , = CONTRAST
##

```

```

##          conc cond confr conj conjalt corr disjalt equiv exempl
## advice     32   0   11   0     0   11     0     0     0
## other      29   0    9   0     0  15     0     0     0
## collection 17   0   15   0     0   8     0     0     0
## comment    77   0   57   0     0  36     0     0     0

```

##	description	84	0	92	0	0	39	0	0	0
##	essay	144	0	109	0	0	123	0	0	0
##	invitation	5	0	4	0	0	0	0	0	0
##	letter	13	0	4	0	0	8	0	0	0
##	news	200	0	176	0	0	83	0	0	0
##	overview	5	0	10	0	0	1	0	0	0
##	person_interv	30	0	23	0	0	24	0	0	0
##	review	61	0	32	0	0	33	0	0	0
##	sport	133	0	72	0	0	29	0	0	0
##	survey	3	0	9	0	0	3	0	0	0
##	topic_interv	47	0	30	0	0	32	0	0	0

##		explicitat	f_cond	f_opp	f_reason	gener	grad	opp	other
##	advice	0	0	1	0	0	14	99	0
##	other	0	0	2	0	0	12	63	0
##	collection	0	0	1	0	0	6	77	0
##	comment	0	0	5	0	0	54	273	0
##	description	0	0	6	0	0	51	328	0
##	essay	0	0	7	0	0	96	562	0
##	invitation	0	0	0	0	0	4	38	0
##	letter	0	0	0	0	0	4	41	0
##	news	0	0	9	0	0	87	673	0
##	overview	0	0	0	0	0	0	18	0
##	person_interv	0	0	4	0	0	14	129	0
##	review	0	0	2	0	0	40	169	0
##	sport	0	0	7	0	0	39	486	0
##	survey	0	0	3	0	0	2	16	0
##	topic_interv	0	0	3	0	0	22	224	0

##		preced	purp	reason	restr	spec	synchr
##	advice	0	0	0	14	0	0
##	other	0	0	0	7	0	0
##	collection	0	0	0	6	0	0
##	comment	0	0	0	21	0	0
##	description	0	0	0	51	0	0
##	essay	0	0	0	51	0	0
##	invitation	0	0	0	1	0	0
##	letter	0	0	0	0	0	0
##	news	0	0	0	61	0	0
##	overview	0	0	0	1	0	0
##	person_interv	0	0	0	3	0	0
##	review	0	0	0	9	0	0
##	sport	0	0	0	28	0	0
##	survey	0	0	0	0	0	0
##	topic_interv	0	0	0	16	0	0

## , , = EXPANSION

##	conc	cond	confr	conj	conjalt	corr	disjalt	equiv	exempl
----	------	------	-------	------	---------	------	---------	-------	--------

##	advice	0	0	0	228	10	0	19	10	8
##	other	0	0	0	265	0	0	10	1	0
##	collection	0	0	0	188	3	0	4	2	5
##	comment	0	0	0	534	6	0	23	15	8
##	description	0	0	0	857	11	0	38	11	37
##	essay	0	0	0	1310	21	0	63	15	24
##	invitation	0	0	0	107	5	0	3	2	1
##	letter	0	0	0	93	4	0	4	0	1
##	news	0	0	0	1762	8	0	51	24	27
##	overview	0	0	0	65	1	0	3	0	0
##	person_interv	0	0	0	266	3	0	7	4	8
##	review	0	0	0	401	10	0	7	2	5
##	sport	0	0	0	936	2	0	10	7	7
##	survey	0	0	0	61	1	0	2	1	0
##	topic_interv	0	0	0	425	5	0	27	11	17

##		explicitat	f_cond	f_opp	f_reason	gener	grad	opp	other
##	advice	0	0	0	0	7	0	0	0
##	other	0	0	0	0	4	0	0	0
##	collection	0	0	0	0	4	0	0	0
##	comment	0	0	0	0	15	0	0	1
##	description	0	0	0	0	6	0	0	0
##	essay	0	0	0	0	25	0	0	1
##	invitation	0	0	0	0	0	0	0	0
##	letter	0	0	0	0	1	0	0	0
##	news	0	0	0	0	17	0	0	0
##	overview	0	0	0	0	0	0	0	0
##	person_interv	0	0	0	0	3	0	0	0
##	review	0	0	0	0	7	0	0	0
##	sport	0	0	0	0	5	0	0	0
##	survey	0	0	0	0	1	0	0	0
##	topic_interv	0	0	0	0	11	0	0	0

##		preced	purp	reason	restr	spec	synchr
##	advice	0	0	0	0	11	0
##	other	0	0	0	0	28	0
##	collection	0	0	0	0	19	0
##	comment	0	0	0	0	54	0
##	description	0	0	0	0	74	0
##	essay	0	0	0	0	110	0
##	invitation	0	0	0	0	10	0
##	letter	0	0	0	0	8	0
##	news	0	0	0	0	90	0
##	overview	0	0	0	0	9	0
##	person_interv	0	0	0	0	15	0
##	review	0	0	0	0	50	0
##	sport	0	0	0	0	128	0
##	survey	0	0	0	0	2	0
##	topic_interv	0	0	0	0	22	0

##

```

## , , = TEMPORAL
##
##
##      conc cond confr conj conjalt corr disjalt equiv exempl
## advice      0  0   0  0      0  0      0  0      0
## other       0  0   0  0      0  0      0  0      0
## collection  0  0   0  0      0  0      0  0      0
## comment     0  0   0  0      0  0      0  0      0
## description 0  0   0  0      0  0      0  0      0
## essay       0  0   0  0      0  0      0  0      0
## invitation  0  0   0  0      0  0      0  0      0
## letter      0  0   0  0      0  0      0  0      0
## news       0  0   0  0      0  0      0  0      0
## overview   0  0   0  0      0  0      0  0      0
## person_interv 0  0   0  0      0  0      0  0      0
## review     0  0   0  0      0  0      0  0      0
## sport      0  0   0  0      0  0      0  0      0
## survey     0  0   0  0      0  0      0  0      0
## topic_interv 0  0   0  0      0  0      0  0      0
##
##      explicat f_cond f_opp f_reason gener grad opp other
## advice      0    0    0    0    0    0    0    0
## other       0    0    0    0    0    0    0    0
## collection  0    0    0    0    0    0    0    0
## comment     0    0    0    0    0    0    0    0
## description 0    0    0    0    0    0    0    0
## essay       0    0    0    0    0    0    0    0
## invitation  0    0    0    0    0    0    0    0
## letter      0    0    0    0    0    0    0    0
## news       0    0    0    0    0    0    0    0
## overview   0    0    0    0    0    0    0    0
## person_interv 0    0    0    0    0    0    0    0
## review     0    0    0    0    0    0    0    0
## sport      0    0    0    0    0    0    0    0
## survey     0    0    0    0    0    0    0    0
## topic_interv 0    0    0    0    0    0    0    0
##
##      preced purp reason restr spec synchr
## advice      21  0    0    0    0    3
## other       39  0    0    0    0   16
## collection  32  0    0    0    0    5
## comment     50  0    0    0    0   19
## description 80  0    0    0    0   20
## essay      142  0    0    0    0   41
## invitation   5  0    0    0    0    2
## letter       6  0    0    0    0    2
## news      199  0    0    0    0   60
## overview    3  0    0    0    0    0
## person_interv 50  0    0    0    0   11
## review     23  0    0    0    0   12

```

```
## sport          147    0    0    0    0    25
## survey         4     0    0    0    0    0
## topic_interv  39     0    0    0    0   10
```

Cross-tabulate genres vs. discourse\_type, by discourse\_class, use xtabs

```
xtabs(~ pdt30$genre + pdt30$discourse_type + pdt30$discourse_class)

## , , pdt30$discourse_class = CONTINGENCY
##
##          pdt30$discourse_type
## pdt30$genre conc cond confr conj conjalt corr disjalt equiv exempl
## advice      0 102   0  0      0  0      0  0      0  0
## other       0  31   0  0      0  0      0  0      0  0
## collection  0  24   0  0      0  0      0  0      0  0
## comment     0 149   0  0      0  0      0  0      0  0
## description 0 156   0  0      0  0      0  0      0  0
## essay       0 278   0  0      0  0      0  0      0  0
## invitation  0  7    0  0      0  0      0  0      0  0
## letter      0  31   0  0      0  0      0  0      0  0
## news        0 231   0  0      0  0      0  0      0  0
## overview    0  10   0  0      0  0      0  0      0  0
## person_interv 0  61   0  0      0  0      0  0      0  0
## review      0  51   0  0      0  0      0  0      0  0
## sport       0 112   0  0      0  0      0  0      0  0
## survey      0  14   0  0      0  0      0  0      0  0
## topic_interv 0 112   0  0      0  0      0  0      0  0
##
##          pdt30$discourse_type
## pdt30$genre explicat f_cond f_opp f_reason gener grad opp other
## advice      6     0     0     2     0     0     0     0
## other       2     0     0     2     0     0     0     0
## collection  9     0     0     0     0     0     0     0
## comment    34     1     0     2     0     0     0     0
## description 17     1     0     2     0     0     0     0
## essay      54     7     0    15     0     0     0     0
## invitation  6     0     0     1     0     0     0     0
## letter     5     0     0     0     0     0     0     0
## news      36     3     0     8     0     0     0     0
## overview   0     0     0     1     0     0     0     0
## person_interv 4     2     0     3     0     0     0     0
## review    23     0     0     1     0     0     0     0
## sport     27     1     0     2     0     0     0     0
## survey     0     0     0     0     0     0     0     0
## topic_interv 7     1     0     1     0     0     0     0
##
##          pdt30$discourse_type
## pdt30$genre preced purp reason restr spec synchr
## advice      0  15  117   0   0   0
## other       0  17   69   0   0   0
## collection  0  17   65   0   0   0
## comment     0  39  258   0   0   0
## description 0  39  305   0   0   0
```

```

##  essay          0  77  482  0  0  0
##  invitation     0   1  16   0  0  0
##  letter        0   6  48   0  0  0
##  news          0  97  608  0  0  0
##  overview      0   0   7   0  0  0
##  person_interv 0  22  81   0  0  0
##  review        0  24  110  0  0  0
##  sport         0  33  289  0  0  0
##  survey        0   1  18   0  0  0
##  topic_interv  0  26  159  0  0  0
##
##  , , pdt30$discourse_class = CONTRAST
##
##                pdt30$discourse_type
## pdt30$genre    conc cond confr conj conjalt corr disjalt equiv exempl
##  advice        32   0   11   0     0  11     0     0     0
##  other         29   0    9   0     0  15     0     0     0
##  collection    17   0   15   0     0   8     0     0     0
##  comment       77   0   57   0     0  36     0     0     0
##  description   84   0   92   0     0  39     0     0     0
##  essay        144   0  109   0     0 123     0     0     0
##  invitation     5   0    4   0     0   0     0     0     0
##  letter        13   0    4   0     0   8     0     0     0
##  news         200   0  176   0     0  83     0     0     0
##  overview      5   0   10   0     0   1     0     0     0
##  person_interv 30   0   23   0     0  24     0     0     0
##  review        61   0   32   0     0  33     0     0     0
##  sport        133   0   72   0     0  29     0     0     0
##  survey         3   0    9   0     0   3     0     0     0
##  topic_interv  47   0   30   0     0  32     0     0     0
##
##                pdt30$discourse_type
## pdt30$genre    explicat f_cond f_opp f_reason gener grad opp other
##  advice         0     0     0     1     0     0  14  99   0
##  other          0     0     0     2     0     0  12  63   0
##  collection     0     0     0     1     0     0   6  77   0
##  comment        0     0     0     5     0     0  54 273   0
##  description    0     0     0     6     0     0  51 328   0
##  essay          0     0     0     7     0     0  96 562   0
##  invitation     0     0     0     0     0     0   4  38   0
##  letter         0     0     0     0     0     0   4  41   0
##  news           0     0     0     9     0     0  87 673   0
##  overview      0     0     0     0     0     0   0  18   0
##  person_interv 0     0     0     4     0     0  14 129   0
##  review         0     0     0     2     0     0  40 169   0
##  sport          0     0     0     7     0     0  39 486   0
##  survey         0     0     0     3     0     0   2  16   0
##  topic_interv  0     0     0     3     0     0  22 224   0
##
##                pdt30$discourse_type
## pdt30$genre    preced purp reason restr spec synchr
##  advice         0     0     0     14   0     0

```



```

## other          0  0  0  7  0  0
## collection    0  0  0  6  0  0
## comment       0  0  0 21  0  0
## description   0  0  0 51  0  0
## essay         0  0  0 51  0  0
## invitation    0  0  0  1  0  0
## letter        0  0  0  0  0  0
## news          0  0  0 61  0  0
## overview      0  0  0  1  0  0
## person_interv 0  0  0  3  0  0
## review        0  0  0  9  0  0
## sport         0  0  0 28  0  0
## survey        0  0  0  0  0  0
## topic_interv  0  0  0 16  0  0
##
## , , pdt30$discourse_class = EXPANSION
##
##           pdt30$discourse_type
## pdt30$genre conc cond confr conj conjalt corr disjalt equiv exempl
## advice      0  0  0 228  10  0  19  10  8
## other       0  0  0 265  0  0  10  1  0
## collection  0  0  0 188  3  0  4  2  5
## comment     0  0  0 534  6  0  23  15  8
## description 0  0  0 857  11 0  38  11 37
## essay       0  0  0 1310 21 0  63  15 24
## invitation  0  0  0 107  5  0  3  2  1
## letter      0  0  0  93  4  0  4  0  1
## news        0  0  0 1762 8  0  51  24 27
## overview    0  0  0  65  1  0  3  0  0
## person_interv 0  0  0 266  3  0  7  4  8
## review      0  0  0 401  10 0  7  2  5
## sport       0  0  0 936  2  0  10  7  7
## survey      0  0  0  61  1  0  2  1  0
## topic_interv 0  0  0 425  5  0  27  11 17
##
##           pdt30$discourse_type
## pdt30$genre explicat f_cond f_opp f_reason gener grad opp other
## advice      0  0  0  0  7  0  0  0
## other       0  0  0  0  4  0  0  0
## collection  0  0  0  0  4  0  0  0
## comment     0  0  0  0  15 0  0  1
## description 0  0  0  0  6  0  0  0
## essay       0  0  0  0  25 0  0  1
## invitation  0  0  0  0  0  0  0  0
## letter      0  0  0  0  1  0  0  0
## news        0  0  0  0  17 0  0  0
## overview    0  0  0  0  0  0  0  0
## person_interv 0  0  0  0  3  0  0  0
## review      0  0  0  0  7  0  0  0
## sport       0  0  0  0  5  0  0  0
## survey      0  0  0  0  1  0  0  0

```

```

## topic_interv      0      0      0      0      11      0      0      0
##                pdt30$discourse_type
## pdt30$genre      preced purp reason restr spec synchr
## advice           0      0      0      0      11      0
## other            0      0      0      0      28      0
## collection       0      0      0      0      19      0
## comment          0      0      0      0      54      0
## description      0      0      0      0      74      0
## essay            0      0      0      0     110      0
## invitation       0      0      0      0      10      0
## letter           0      0      0      0       8      0
## news             0      0      0      0      90      0
## overview         0      0      0      0       9      0
## person_interv   0      0      0      0      15      0
## review           0      0      0      0      50      0
## sport            0      0      0      0     128      0
## survey           0      0      0      0       2      0
## topic_interv    0      0      0      0      22      0
##
## , , pdt30$discourse_class = TEMPORAL
##
##                pdt30$discourse_type
## pdt30$genre      conc cond confr conj conjalt corr disjalt equiv exempl
## advice           0      0      0      0          0      0          0      0      0
## other            0      0      0      0          0      0          0      0      0
## collection       0      0      0      0          0      0          0      0      0
## comment          0      0      0      0          0      0          0      0      0
## description      0      0      0      0          0      0          0      0      0
## essay            0      0      0      0          0      0          0      0      0
## invitation       0      0      0      0          0      0          0      0      0
## letter           0      0      0      0          0      0          0      0      0
## news             0      0      0      0          0      0          0      0      0
## overview         0      0      0      0          0      0          0      0      0
## person_interv   0      0      0      0          0      0          0      0      0
## review           0      0      0      0          0      0          0      0      0
## sport            0      0      0      0          0      0          0      0      0
## survey           0      0      0      0          0      0          0      0      0
## topic_interv    0      0      0      0          0      0          0      0      0
##
##                pdt30$discourse_type
## pdt30$genre      explicat f_cond f_opp f_reason gener grad opp other
## advice           0      0      0          0      0      0      0      0
## other            0      0      0          0      0      0      0      0
## collection       0      0      0          0      0      0      0      0
## comment          0      0      0          0      0      0      0      0
## description      0      0      0          0      0      0      0      0
## essay            0      0      0          0      0      0      0      0
## invitation       0      0      0          0      0      0      0      0
## letter           0      0      0          0      0      0      0      0
## news             0      0      0          0      0      0      0      0
## overview         0      0      0          0      0      0      0      0

```

```

## person_interv      0      0      0      0      0      0      0      0
## review              0      0      0      0      0      0      0      0
## sport               0      0      0      0      0      0      0      0
## survey              0      0      0      0      0      0      0      0
## topic_interv       0      0      0      0      0      0      0      0
##
##                pdt30$discourse_type
## pdt30$genre      preced purp reason restr spec synchr
## advice           21      0        0        0        0        3
## other            39      0        0        0        0       16
## collection       32      0        0        0        0        5
## comment          50      0        0        0        0       19
## description      80      0        0        0        0       20
## essay           142      0        0        0        0       41
## invitation        5      0        0        0        0        2
## letter           6      0        0        0        0        2
## news            199      0        0        0        0       60
## overview         3      0        0        0        0        0
## person_interv    50      0        0        0        0       11
## review           23      0        0        0        0       12
## sport           147      0        0        0        0       25
## survey           4      0        0        0        0        0
## topic_interv     39      0        0        0        0       10

```

table and xtabs do the same thing, but NB the *formula* in xtabs.

Run Pearson's Chi-Squared Test for discourse\_class and genre. Interpret the standardized residuals. To be able to do this, you have to store the test results in a variable (call it results). Look at the structure and access the right element of results.

```

results <- chisq.test(pdt30$genre, pdt30$discourse_class)
str(results)

## List of 9
## $ statistic: Named num 273
## .. attr(*, "names")= chr "X-squared"
## $ parameter: Named int 42
## .. attr(*, "names")= chr "df"
## $ p.value : num 1.1e-35
## $ method : chr "Pearson's Chi-squared test"
## $ data.name: chr "pdt30$genre and pdt30$discourse_class"
## $ observed : 'table' int [1:15, 1:4] 242 121 115 483 520 913 31 90 983 18
...
## .. attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...
## .. ..$ pdt30$discourse_class: chr [1:4] "CONTINGENCY" "CONTRAST"
"EXPANSION" "TEMPORAL"
## $ expected : num [1:15, 1:4] 169 142 116 396 527 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...

```

```

## .. ..$ pdt30$discourse_class: chr [1:4] "CONTINGENCY" "CONTRAST"
"EXPANSION" "TEMPORAL"
## $ residuals: table [1:15, 1:4] 5.5716 -1.7642 -0.0885 4.3783 -0.3119 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...
## .. ..$ pdt30$discourse_class: chr [1:4] "CONTINGENCY" "CONTRAST"
"EXPANSION" "TEMPORAL"
## $ stdres : table [1:15, 1:4] 6.462 -2.04 -0.102 5.21 -0.377 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...
## .. ..$ pdt30$discourse_class: chr [1:4] "CONTINGENCY" "CONTRAST"
"EXPANSION" "TEMPORAL"
## - attr(*, "class")= chr "htest"

```

```
results$stdres
```

```

##          pdt30$discourse_class
## pdt30$genre CONTINGENCY  CONTRAST  EXPANSION  TEMPORAL
## advice      6.4615765 -2.6468625 -1.9676630 -2.4346740
## other       -2.0398666 -3.8116713  3.3434255  4.1887851
## collection  -0.1020027 -1.6335066  0.6094034  2.1712639
## comment     5.2095058  1.2713551 -4.5289432 -2.3526586
## description -0.3769206 -0.7255357  1.8561264 -1.9478011
## essay       2.3100952  0.2653760 -1.7707918 -0.9637207
## invitation  -3.0571787 -1.6490094  4.6946794 -1.3221224
## letter      3.7590890 -1.4095414 -1.1107431 -1.7585667
## news        -1.9443364 -0.5155531  1.2665335  1.9083080
## overview    -2.6096549 -0.7095047  3.5542873 -1.5436120
## person_interv -0.2116568  0.4405930 -1.8010955  3.5221289
## review      -2.7016956  2.5136831  1.2955250 -2.9136409
## sport       -5.7410089  3.0270799  0.3371150  3.9337022
## survey      0.1514688 -0.8823696  1.2445159 -1.2623203
## topic_interv 1.4474858  0.8871888 -1.1151822 -2.0648269

```

Run Pearson's Chi-Squared Test for *discourse\_type* and *genre*. Store the result in *results* again (overwrite the original variable content)

```

results <- chisq.test(pdt30$genre, pdt30$discourse_type)

## Warning in chisq.test(pdt30$genre, pdt30$discourse_type): Chi-squared
## approximation may be incorrect

str(results)

## List of 9
## $ statistic: Named num 1028
## ..- attr(*, "names")= chr "X-squared"
## $ parameter: Named int 308
## ..- attr(*, "names")= chr "df"

```

```

## $ p.value : num 2.22e-78
## $ method : chr "Pearson's Chi-squared test"
## $ data.name: chr "pdt30$genre and pdt30$discourse_type"
## $ observed : 'table' int [1:15, 1:23] 32 29 17 77 84 144 5 13 200 5 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...
## .. ..$ pdt30$discourse_type: chr [1:23] "conc" "cond" "confr" "conj" ...
## $ expected : num [1:15, 1:23] 31.7 26.6 21.7 74.1 98.7 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...
## .. ..$ pdt30$discourse_type: chr [1:23] "conc" "cond" "confr" "conj" ...
## $ residuals: table [1:15, 1:23] 0.0491 0.4681 -1.01 0.336 -1.4779 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...
## .. ..$ pdt30$discourse_type: chr [1:23] "conc" "cond" "confr" "conj" ...
## $ stdres : table [1:15, 1:23] 0.0511 0.4859 -1.0453 0.3589 -1.6032 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ pdt30$genre : chr [1:15] "advice" "other" "collection"
"comment" ...
## .. ..$ pdt30$discourse_type: chr [1:23] "conc" "cond" "confr" "conj" ...
## - attr(*, "class")= chr "htest"

```

results\$stdres

```

##          pdt30$discourse_type
## pdt30$genre      conc      cond      confr      conj
## advice           0.051075566  7.900771794 -2.675393130 -3.287969156
## other            0.485881096 -1.693117806 -2.492612720  3.256846500
## collection      -1.045344084 -1.761606754 -0.283739185  0.285665462
## comment         0.358875596  3.395986576  0.287683674 -5.083569763
## description    -1.603184809  0.220123301  2.366225015  0.743376725
## essay          -1.501716273  2.010568214 -1.065440216 -2.267063587
## invitation     -1.457455875 -2.053438899 -1.135800855  3.886900870
## letter          0.314287020  3.002231034 -1.671541661 -1.098580372
## news            0.575972169 -4.689446893  3.144740428  4.091349076
## overview       -0.315455722  0.373804708  2.838059660  2.902359590
## person_interv -0.515708682  1.463654355 -0.286711457 -1.053822450
## review         2.340934903 -2.566409924 -0.367719673  0.649014999
## sport          2.613632303 -4.786903029 -0.995324922  0.659350862
## survey         -1.267580056  1.562173079  2.178204877  1.679406960
## topic_interv  -0.921807398  3.392284739 -1.601876240 -1.813524674
##          pdt30$discourse_type
## pdt30$genre      conjalt      corr      disjalt      equiv
## advice           3.828367220 -1.296298208  3.027940058  3.261866347
## other            -1.678118766  0.435620322  0.647568122 -1.241640503
## collection       0.531326001 -0.919617382 -1.058317242 -0.372100005
## comment         -0.600724903 -0.254522266  0.039250778  2.169420421

```

```

## description 0.303850507 -1.655803882 1.474905929 -0.240156087
## essay 1.243683699 5.166542097 2.130797776 -1.061142574
## invitation 4.172038314 -2.208093774 0.075131495 0.846630902
## letter 2.536591475 0.811750986 0.169969583 -1.205085038
## news -2.998600962 -1.695177212 -1.250118641 0.227387178
## overview 0.542505434 -1.132089558 0.937133553 -0.832186431
## person_interv -0.199712721 1.869958381 -1.004126873 0.042317374
## review 2.521225068 2.110752086 -1.961940965 -1.529658062
## sport -2.914357107 -3.747065697 -4.338652303 -1.758064514
## survey 0.489730020 -0.030508640 0.104489319 0.331597821
## topic_interv -0.203559452 1.004509412 2.704899980 1.897647892
## pdt30$discourse_type
## pdt30$genre exempl explicit f_cond f_opp
## advice 1.179238682 -0.815089445 -0.773841665 -0.609538284
## other -2.155006523 -1.917091625 -0.706281218 0.404854346
## collection 0.717777094 1.422331703 -0.636352182 -0.212957508
## comment -1.325959977 3.493544260 -0.312869799 0.402461245
## description 5.334511429 -1.847595298 -0.629499504 0.176395830
## essay -0.651279287 2.052192219 2.637325804 -0.783612839
## invitation -0.458760921 2.304939876 -0.414299002 -0.732991027
## letter -0.719268408 1.076291373 -0.469396714 -0.830471659
## news -1.090971459 -2.317637653 -0.308555586 -0.674208706
## overview -0.989040128 -1.235438918 -0.324147727 -0.573492513
## person_interv 1.078356045 -1.603192637 1.851294907 1.594457153
## review -1.008735168 3.282134980 -0.938638928 -0.386973283
## sport -2.809882229 -0.253210166 -0.735582944 0.370075863
## survey -1.014718357 -1.267514343 -0.332563503 4.558082463
## topic_interv 2.772054558 -1.931460694 0.030734386 -0.019766626
## pdt30$discourse_type
## pdt30$genre f_reason gener grad opp
## advice 0.473763086 1.660531317 -0.525010781 -1.674254237
## other 0.731876007 0.453744814 -0.404358639 -3.773482284
## collection -1.006749514 0.869840750 -1.537606275 -0.227231480
## comment -0.779934669 2.129685925 2.851883581 0.267194490
## description -1.246700081 -1.816721022 0.166851355 -1.854066972
## essay 3.148649792 1.417431819 1.818394776 -1.103559532
## invitation 0.889589310 -1.068711391 -0.336642686 0.771202058
## letter -0.742615374 -0.369262252 -0.844998823 -0.395973480
## news -0.296939340 -1.472438032 -1.231952158 -1.313396778
## overview 1.453793714 -0.836160278 -1.727613572 -0.678073749
## person_interv 1.258768458 -0.490847155 -0.658708079 0.989452815
## review -0.773141782 0.644633558 3.619660614 0.200831239
## sport -1.404813729 -2.379556518 -2.287036213 5.476804302
## survey -0.526136555 0.321944936 -0.611161499 -1.381369186
## topic_interv -0.945864951 1.863974019 -1.003142002 2.427824156
## pdt30$discourse_type
## pdt30$genre other preced purp reason
## advice -0.273501147 -1.754162207 0.020097159 2.476661573
## other -0.249623059 2.803776984 1.303075567 -1.282557676
## collection -0.224907832 2.562385641 2.173041049 0.010809809

```

```

## comment      2.117483004 -2.631003106  0.739415736  2.732383508
## description  -0.502620454 -1.585035871 -1.168360411  0.651779569
## essay        1.160827373 -1.051351456  0.170851786  0.050150582
## invitation   -0.146426921 -1.344336130 -1.643464960 -2.427842146
## letter       -0.165900268 -1.644628469  0.163325371  2.214395085
## news         -0.749815714  1.250933509  0.739468025  1.538873658
## overview    -0.114564489 -1.083939203 -1.665069221 -2.634852316
## person_interv -0.278440791  3.467795410  1.716313950 -1.895641198
## review       -0.331745931 -3.297234840  0.537902072 -2.559912440
## sport        -0.529259857  4.702370686 -2.700828331 -2.182330010
## survey       -0.117538902 -0.752134554 -1.106762625 -0.013804786
## topic_interv -0.359418733 -1.765072503  0.183873402 -0.058957691
##
## pdt30$discourse_type
## pdt30$genre      restr      spec      synchr
## advice           1.416585693 -2.542232772 -1.846890308
## other            -0.404068429  2.119801391  3.584207694
## collection       -0.251278437  0.902886439 -0.247711729
## comment          -0.365388096  0.137810107 -0.007745393
## description      4.052518475  0.429999223 -1.132662696
## essay            0.291040670 -0.539227217 -0.053271504
## invitation       -1.110169943  1.310888215 -0.259154362
## letter           -1.936634349 -0.192764017 -0.617119209
## news             0.293391167 -4.715830425  1.682996930
## overview         -0.574742009  2.460302116 -1.224528363
## person_interv   -2.278915454 -1.816532738  0.905893954
## review           -1.388245269  3.120236900  0.064222591
## sport            -0.943130451  6.239002648 -0.562827231
## survey           -1.372088648 -1.138193561 -1.256320520
## topic_interv    -0.082088992 -2.749495571 -1.039704146

```

There were some expected values lower than 5, but no zeros. The test finished with meaningful residuals at least. Let's look where we had 5 and less in expected values.

## Accessing a Matrix, Matrix Transformation, Pipe Operator

```

which(results$expected < 5, arr.ind = TRUE) %>% row.names() %>% table() %>%
prop.table()*100

```

```

## .
##      advice  collection  comment  description  essay
##      6.140351  7.017544  3.508772  2.631579  1.754386
##      invitation  letter      news      other      overview
##      13.157895  10.526316  1.754386  7.017544  14.912281
## person_interv  review      sport      survey  topic_interv
##      6.140351  4.385965  2.631579  14.912281  3.508772

```

```

which(t(results$expected) < 5, arr.ind = TRUE) %>% row.names() %>% table()
%>% prop.table()*100

```

```

## .
##      confr  conjalt  corr  disjalt  equiv  exempl  explicat

```

```
## 1.754386 7.894737 2.631579 3.508772 7.017544 5.263158 3.508772
## f_cond f_opp f_reason gener grad other purp
## 13.157895 9.649123 11.403509 7.017544 2.631579 13.157895 2.631579
## restr spec synchr
## 3.508772 1.754386 3.508772
```

We have decided that we believe the results. Let's analyze the standardized residuals. Just look:

```
results$stdres

##          pdt30$discourse_type
## pdt30$genre      conc      cond      confr      conj
## advice          0.051075566  7.900771794 -2.675393130 -3.287969156
## other           0.485881096 -1.693117806 -2.492612720  3.256846500
## collection     -1.045344084 -1.761606754 -0.283739185  0.285665462
## comment        0.358875596  3.395986576  0.287683674 -5.083569763
## description    -1.603184809  0.220123301  2.366225015  0.743376725
## essay         -1.501716273  2.010568214 -1.065440216 -2.267063587
## invitation     -1.457455875 -2.053438899 -1.135800855  3.886900870
## letter         0.314287020  3.002231034 -1.671541661 -1.098580372
## news           0.575972169 -4.689446893  3.144740428  4.091349076
## overview      -0.315455722  0.373804708  2.838059660  2.902359590
## person_interv -0.515708682  1.463654355 -0.286711457 -1.053822450
## review         2.340934903 -2.566409924 -0.367719673  0.649014999
## sport         2.613632303 -4.786903029 -0.995324922  0.659350862
## survey        -1.267580056  1.562173079  2.178204877  1.679406960
## topic_interv  -0.921807398  3.392284739 -1.601876240 -1.813524674
##          pdt30$discourse_type
## pdt30$genre      conjalt      corr      disjalt      equiv
## advice          3.828367220 -1.296298208  3.027940058  3.261866347
## other           -1.678118766  0.435620322  0.647568122 -1.241640503
## collection     0.531326001 -0.919617382 -1.058317242 -0.372100005
## comment        -0.600724903 -0.254522266  0.039250778  2.169420421
## description    0.303850507 -1.655803882  1.474905929 -0.240156087
## essay         1.243683699  5.166542097  2.130797776 -1.061142574
## invitation     4.172038314 -2.208093774  0.075131495  0.846630902
## letter         2.536591475  0.811750986  0.169969583 -1.205085038
## news          -2.998600962 -1.695177212 -1.250118641  0.227387178
## overview      0.542505434 -1.132089558  0.937133553 -0.832186431
## person_interv -0.199712721  1.869958381 -1.004126873  0.042317374
## review         2.521225068  2.110752086 -1.961940965 -1.529658062
## sport         -2.914357107 -3.747065697 -4.338652303 -1.758064514
## survey         0.489730020 -0.030508640  0.104489319  0.331597821
## topic_interv  -0.203559452  1.004509412  2.704899980  1.897647892
##          pdt30$discourse_type
## pdt30$genre      exempl      explicat      f_cond      f_opp
## advice          1.179238682 -0.815089445 -0.773841665 -0.609538284
## other           -2.155006523 -1.917091625 -0.706281218  0.404854346
## collection     0.717777094  1.422331703 -0.636352182 -0.212957508
## comment        -1.325959977  3.493544260 -0.312869799  0.402461245
```



```

## description 5.334511429 -1.847595298 -0.629499504 0.176395830
## essay -0.651279287 2.052192219 2.637325804 -0.783612839
## invitation -0.458760921 2.304939876 -0.414299002 -0.732991027
## letter -0.719268408 1.076291373 -0.469396714 -0.830471659
## news -1.090971459 -2.317637653 -0.308555586 -0.674208706
## overview -0.989040128 -1.235438918 -0.324147727 -0.573492513
## person_interv 1.078356045 -1.603192637 1.851294907 1.594457153
## review -1.008735168 3.282134980 -0.938638928 -0.386973283
## sport -2.809882229 -0.253210166 -0.735582944 0.370075863
## survey -1.014718357 -1.267514343 -0.332563503 4.558082463
## topic_interv 2.772054558 -1.931460694 0.030734386 -0.019766626
## pdt30$discourse_type
## pdt30$genre f_reason gener grad opp
## advice 0.473763086 1.660531317 -0.525010781 -1.674254237
## other 0.731876007 0.453744814 -0.404358639 -3.773482284
## collection -1.006749514 0.869840750 -1.537606275 -0.227231480
## comment -0.779934669 2.129685925 2.851883581 0.267194490
## description -1.246700081 -1.816721022 0.166851355 -1.854066972
## essay 3.148649792 1.417431819 1.818394776 -1.103559532
## invitation 0.889589310 -1.068711391 -0.336642686 0.771202058
## letter -0.742615374 -0.369262252 -0.844998823 -0.395973480
## news -0.296939340 -1.472438032 -1.231952158 -1.313396778
## overview 1.453793714 -0.836160278 -1.727613572 -0.678073749
## person_interv 1.258768458 -0.490847155 -0.658708079 0.989452815
## review -0.773141782 0.644633558 3.619660614 0.200831239
## sport -1.404813729 -2.379556518 -2.287036213 5.476804302
## survey -0.526136555 0.321944936 -0.611161499 -1.381369186
## topic_interv -0.945864951 1.863974019 -1.003142002 2.427824156
## pdt30$discourse_type
## pdt30$genre other preced purp reason
## advice -0.273501147 -1.754162207 0.020097159 2.476661573
## other -0.249623059 2.803776984 1.303075567 -1.282557676
## collection -0.224907832 2.562385641 2.173041049 0.010809809
## comment 2.117483004 -2.631003106 0.739415736 2.732383508
## description -0.502620454 -1.585035871 -1.168360411 0.651779569
## essay 1.160827373 -1.051351456 0.170851786 0.050150582
## invitation -0.146426921 -1.344336130 -1.643464960 -2.427842146
## letter -0.165900268 -1.644628469 0.163325371 2.214395085
## news -0.749815714 1.250933509 0.739468025 1.538873658
## overview -0.114564489 -1.083939203 -1.665069221 -2.634852316
## person_interv -0.278440791 3.467795410 1.716313950 -1.895641198
## review -0.331745931 -3.297234840 0.537902072 -2.559912440
## sport -0.529259857 4.702370686 -2.700828331 -2.182330010
## survey -0.117538902 -0.752134554 -1.106762625 -0.013804786
## topic_interv -0.359418733 -1.765072503 0.183873402 -0.058957691
## pdt30$discourse_type
## pdt30$genre restr spec synchr
## advice 1.416585693 -2.542232772 -1.846890308
## other -0.404068429 2.119801391 3.584207694
## collection -0.251278437 0.902886439 -0.247711729

```

```
## comment      -0.365388096  0.137810107 -0.007745393
## description  4.052518475  0.429999223 -1.132662696
## essay        0.291040670 -0.539227217 -0.053271504
## invitation   -1.110169943  1.310888215 -0.259154362
## letter       -1.936634349 -0.192764017 -0.617119209
## news         0.293391167 -4.715830425  1.682996930
## overview     -0.574742009  2.460302116 -1.224528363
## person_interv -2.278915454 -1.816532738  0.905893954
## review       -1.388245269  3.120236900  0.064222591
## sport        -0.943130451  6.239002648 -0.562827231
## survey       -1.372088648 -1.138193561 -1.256320520
## topic_interv -0.082088992 -2.749495571 -1.039704146
```

It's already a pretty big matrix. Let's find out which discourse type is most attracted to or repelled from a genre:

```
max(abs(results$stdres))

## [1] 7.900772

which(abs(results$stdres) > 7 , arr.ind = TRUE)

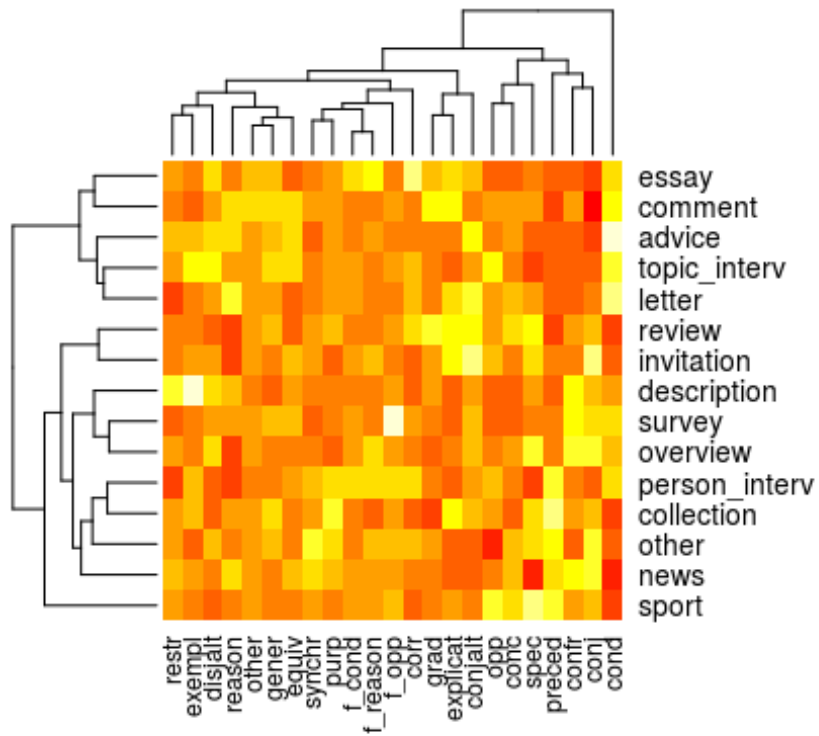
##      pdt30$genre pdt30$discourse_type
## advice          1                    2

which(abs(t(results$stdres)) > 7, arr.ind = TRUE)

##      pdt30$discourse_type pdt30$genre
## cond                      2          1
```

Let's visualize it with a heatmap. The darker the color (here red), the lower the cell value, increasing over yellow to white. Use heatmap, base package.

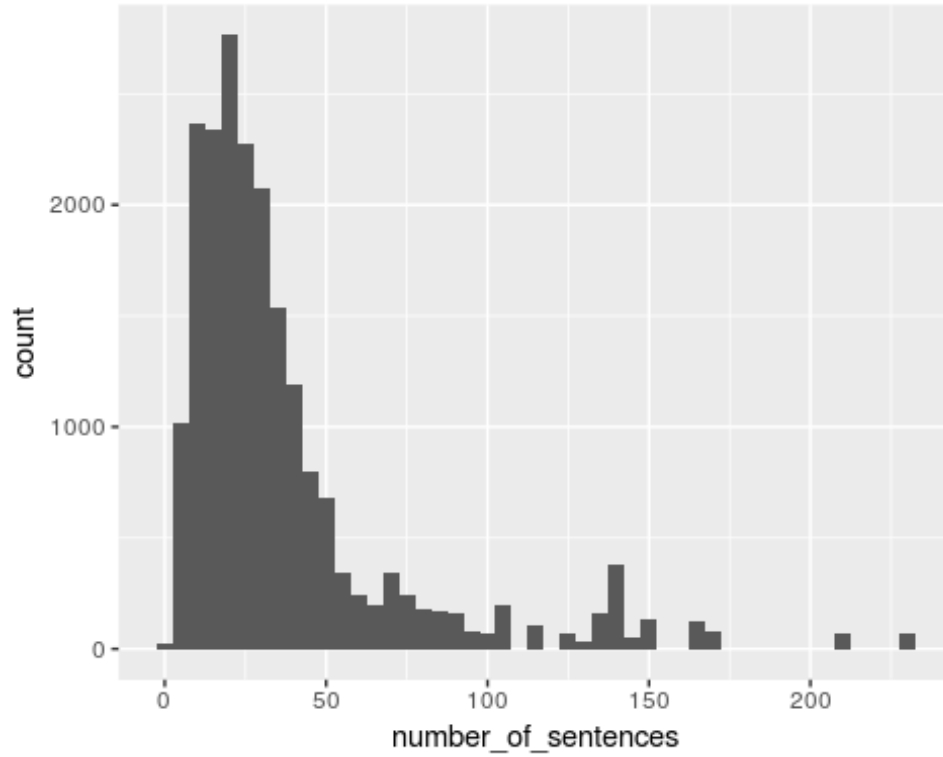
```
heatmap(results$stdres)
```



## Simple Plotting with `ggplot2`

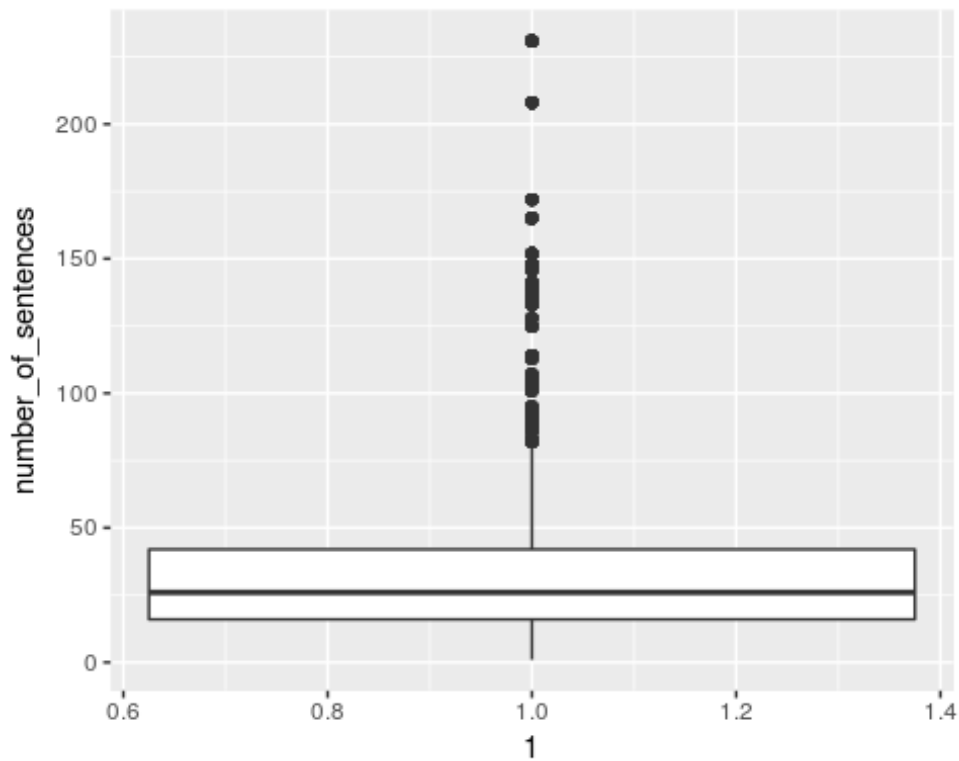
Load the `ggplot2` library. Draw a histogram of sentence length in `pdt30`. Play with `binwidth`.

```
ggplot(data = pdt30, aes(x = number_of_sentences)) + geom_histogram(binwidth = 5)
```



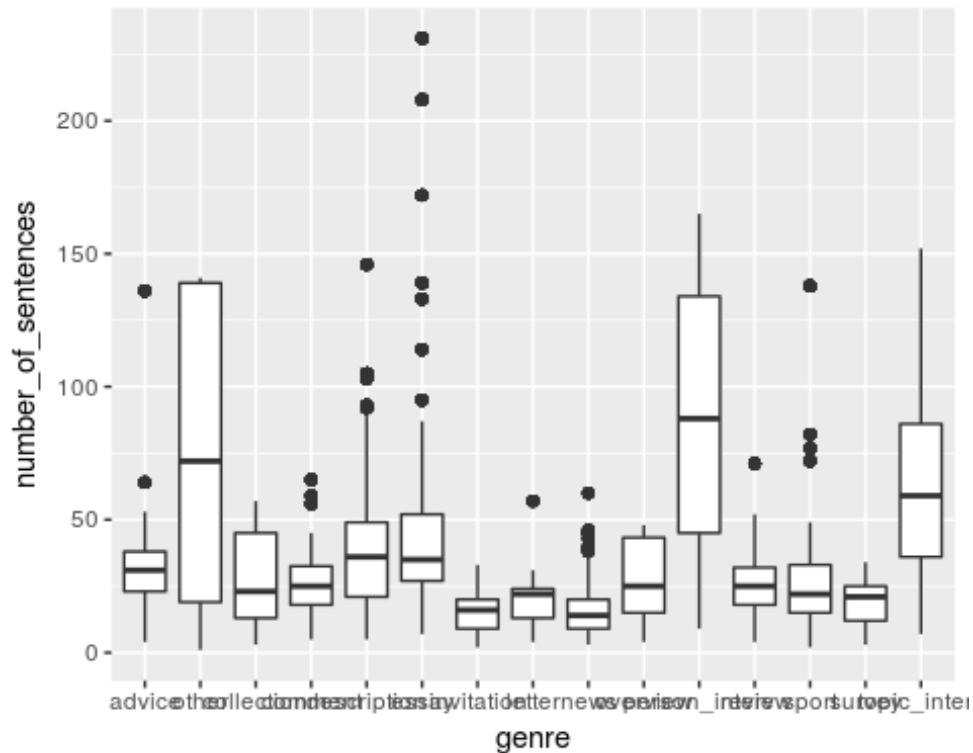
Draw a boxplot of number\_of\_sentences. (Just 1 on x-axis)

```
ggplot(data = pdt30, aes(x = 1, y = number_of_sentences)) + geom_boxplot()
```



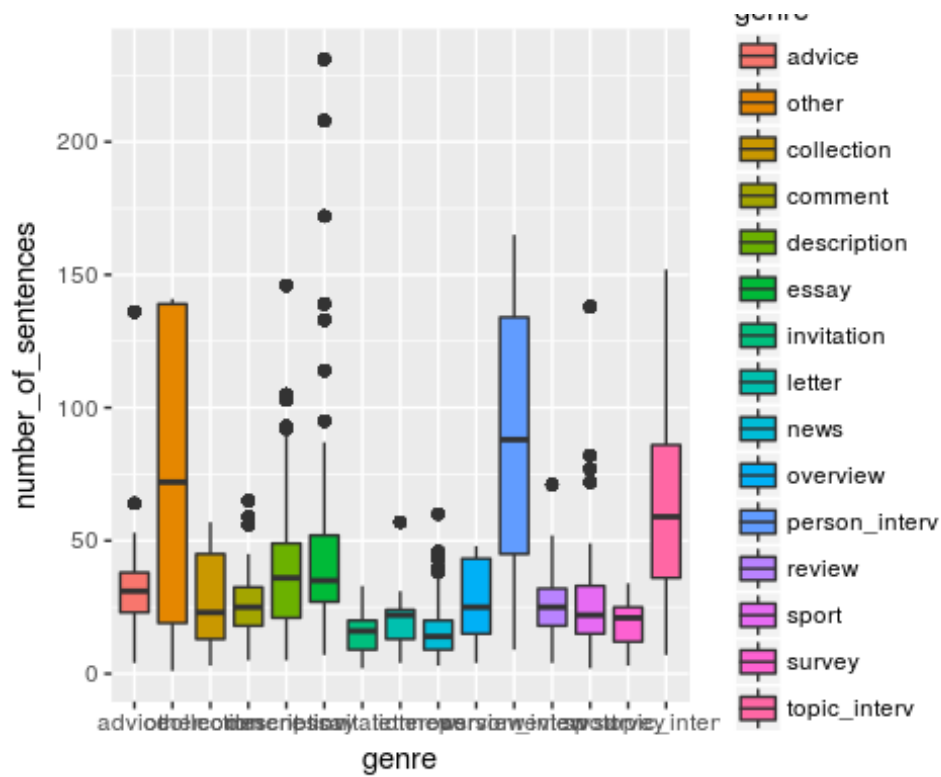
Draw a row of boxplots of number\_of\_sentences according to genre. (Genres on x-axis)

```
ggplot(data = pdt30, aes(x = genre, y = number_of_sentences)) +  
geom_boxplot()
```



Draw a row of boxplots of number\_of\_sentences according to genre, split the boxplots according to discourse\_class using color (here fill).

```
ggplot(data = pdt30, aes(x = genre, y = number_of_sentences, fill = genre)) +  
geom_boxplot()
```

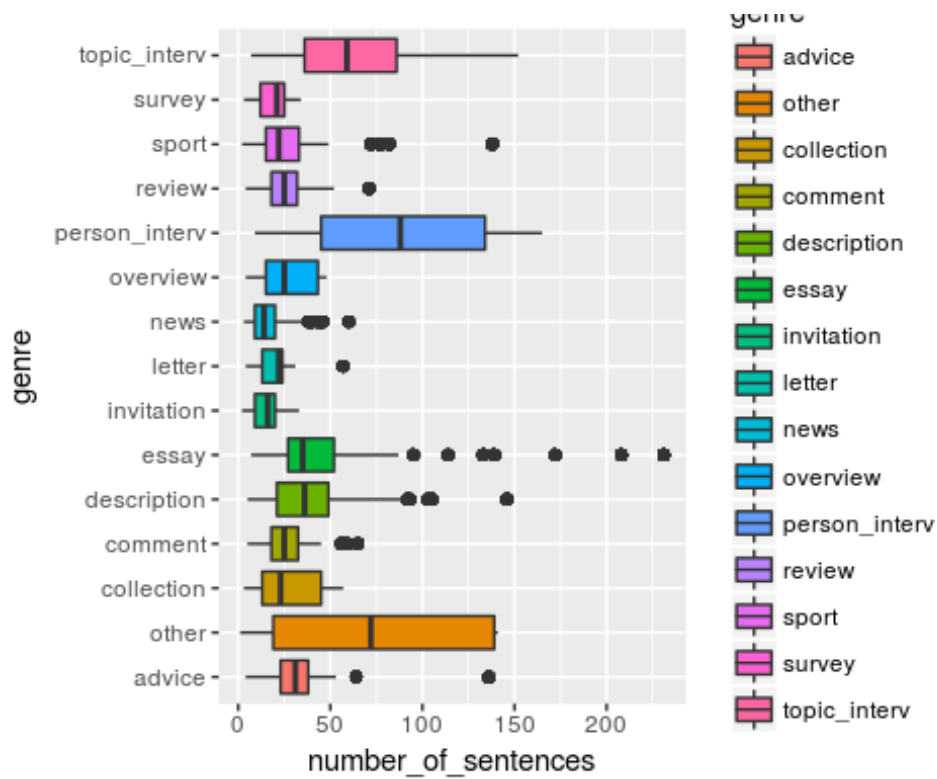


Store this plot in boxplots variable.

```
boxplots <- ggplot(data = pdt30, aes(x = genre, y = number_of_sentences, fill = genre)) + geom_boxplot()
```

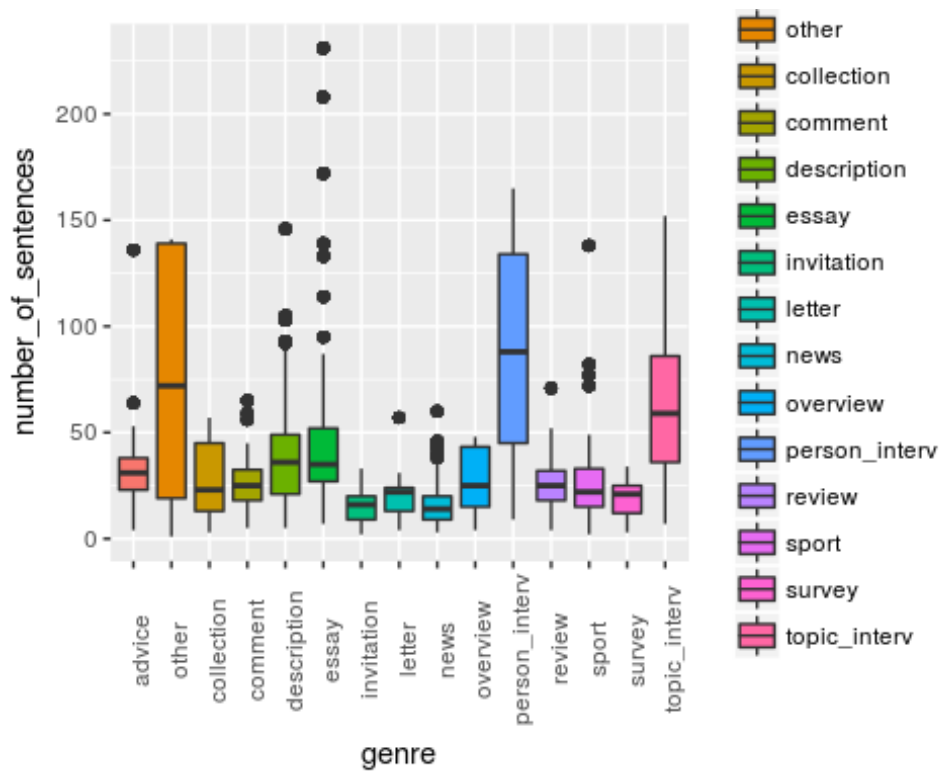
Flip the plot to make it narrower.

```
boxplots + coord_flip()
```



Or leave it wide as it is but flip the ticks on the x-axis by 90 degrees.

boxplots + `theme(axis.text.x = element_text(angle = 90))`



## Export or Save Data

Export in `pdt30` `pdt30_cleaned.csv` in your home folder. (You are not entitled to write the `edu` folder)

```
write.csv(pdt30, file = "pdt30_cleaned.csv")
```

## Save inside R as an Object

`.RData`

```
save(pdt30, file = "pdt30_cleaned.RData")
```

`.RDS`

```
saveRDS(pdt30, file = "pdt30_cleaned.RDS")
```