Datafest

Kemon Lardas

2024-04-12

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.5.0     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
getwd()
```

```
## [1] "C:/Users/13475/OneDrive/Documents/Side Stuff/2024 ASA DataFest Data 03-04/Full Data"
```

```
checkpoints_eoc <- read_csv("checkpoints_eoc.csv")
```

```
## Rows: 16418 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (3): student_id, class_id, book
## dbl (5): chapter_number, EOC, n_possible, n_correct, n_attempt
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
checkpoints_pulse <- read_csv("checkpoints_pulse.csv")
```

```
## Rows: 76848 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (6): book, release, institution_id, class_id, student_id, construct
## dbl (2): chapter_number, response
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
items <- read_csv("items.csv")
```

```
## Rows: 1335 Columns: 19
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (13): institution_id, class_id, item_id, item_type, chapter, page, dcl_s...
## dbl  (3): lrn_question_position, chapter_number, section_number
## lgl  (3): dcl_pre_exercise_code, dcl_hint, review_flag
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
media_views <- read_csv("media_views.csv")
```

```
## Rows: 6149 Columns: 16
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (7): book, release, chapter, page, institution_id, class_id, student_id
## dbl  (6): chapter_number, section_number, media_id, access_count, proportion...
## lgl  (1): review_flag
## dttm (2): dt_started, dt_last_event
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
page_views <- read_csv("page_views.csv")
```

```
## Rows: 478752 Columns: 19
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (7): book, release, chapter, page, institution_id, class_id, student_id
## dbl  (8): chapter_number, section_number, tried_again_clicks, engaged, idle_...
## lgl  (2): was_complete, review_flag
## dttm (2): dt_accessed, tried_again_dt
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
responses <- read_csv("responses.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 208229 Columns: 40
## ── Column specification ────────────────────────────────────────────────────
## Delimiter: ","
## chr  (25): book, release, chapter, page, institution_id, class_id, student_i...
## dbl   (6): chapter_number, section_number, points_possible, points_earned, a...
## lgl   (6): review_flag, completes_page, lrn_option_8, lrn_option_9, lrn_opti...
## dttm  (3): dt_submitted, lrn_dt_started, lrn_dt_saved
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Questions

#Which videos had the least proportion of amount of watches?
prop_watch_amount <- media_views %>%
    group_by(media_id) %>%
    summarize(not_seen = sum(is.na(proportion_video))/length(proportion_video)) %>%
    arrange((not_seen))

#Which media had the most proportion of time watched?
prop_time_watched <- media_views %>%
    na.omit() %>%
    group_by(media_id) %>%
    summarize(prop_time_watched = sum(proportion_video)/length(proportion_video))


#If it was a review video, was the proportion of time watched longer and did more people view th
em?

media_views %>%
    filter(review_flag == FALSE) #no review videos
```

```
## # A tibble: 6,149 × 16
##    book       release chapter page  chapter_number section_number institution_id
##    <chr>      <chr>   <chr>   <chr>          <dbl>          <dbl> <chr>
##  1 College /… v5.0    Chapte… 5.4 …              5              4 04157183-8665…
##  2 College /… v5.0    Chapte… 5.9 …              5              9 04157183-8665…
##  3 College /… v5.0    Chapte… 5.9 …              5              9 04157183-8665…
##  4 College /… v5.0    Chapte… 5.4 …              5              4 04157183-8665…
##  5 College /… v5.0    Chapte… 5.9 …              5              9 04157183-8665…
##  6 College /… v5.0    Chapte… 5.9 …              5              9 04157183-8665…
##  7 College /… v5.0    Chapte… 5.4 …              5              4 04157183-8665…
##  8 College /… v5.0    Chapte… 5.9 …              5              9 04157183-8665…
##  9 College /… v5.0    Chapte… 5.9 …              5              9 04157183-8665…
## 10 College /… v5.0    Chapte… 5.4 …              5              4 04157183-8665…
## # ℹ 6,139 more rows
## # ℹ 9 more variables: class_id <chr>, student_id <chr>, media_id <dbl>,
## #   dt_started <dttm>, dt_last_event <dttm>, access_count <dbl>,
## #   proportion_video <dbl>, proportion_time <dbl>, review_flag <lgl>
```

```
#What was the average time for each video
avg_video_time <- media_views %>%
    na.omit() %>%
    mutate(video_spent_time = dt_last_event - dt_started) %>%
    group_by(media_id) %>%
    summarize(avg_time_spent = mean(video_spent_time))

media_views %>%
    na.omit() %>%
    count(media_id)
```

```
## # A tibble: 5 × 2
##    media_id     n
##       <dbl> <int>
## 1 379060892   340
## 2 379150092   186
## 3 379319375   419
## 4 379319558   210
## 5 381974697   289
```

```
avg_video_time %>%
    inner_join(prop_watch_amount) %>%
    inner_join(prop_time_watched)
```

```
## Joining with `by = join_by(media_id)`
## Joining with `by = join_by(media_id)`
```

```
## # A tibble: 5 × 4
##    media_id avg_time_spent not_seen prop_time_watched
##       <dbl> <drtn>            <dbl>             <dbl>
## 1 379060892 65011.453 secs    0.728             0.741
## 2 379150092 47329.989 secs    0.847             0.753
## 3 379319375 32036.413 secs    0.665             0.777
## 4 379319558  2643.757 secs    0.827             0.784
## 5 381974697 42550.131 secs    0.763             0.655
```

```
#What section had the most proportion of time watched? The middle of the chapter has least engag
ement
media_views %>%
    na.omit() %>%
    group_by(section_number) %>%
    summarize(prop_time_watched = sum(proportion_video)/length(proportion_video))
```

```
## # A tibble: 3 × 2
##   section_number prop_time_watched
##            <dbl>             <dbl>
## 1              2             0.761
## 2              4             0.655
## 3              9             0.769
```

```
#As you go through each chapter less and less students are watching the videos
media_views %>%
    group_by(section_number) %>%
    summarize(not_seen = sum(is.na(proportion_video))/length(proportion_video)) %>%
    arrange((not_seen))
```

```
## # A tibble: 3 × 2
##   section_number not_seen
##            <dbl>    <dbl>
## 1              2    0.696
## 2              4    0.763
## 3              9    0.837
```

```
media_views %>%
    group_by(release) %>%
    summarize(not_seen = sum(is.na(proportion_video))/length(proportion_video))
```

```
## # A tibble: 5 × 2
##   release   not_seen
##   <chr>        <dbl>
## 1 v5.0         0.779
## 2 v5.0-exp1    0.742
## 3 v5.0-exp2    0.738
## 4 v5.1.1       0.792
## 5 v5.2         0.763
```

```
#media_views %>%
#    na.omit() %>%
#    mutate(time_started = as.ITime(dt_started))

media_views %>%
    na.omit() %>%
    mutate(hours = hour(dt_started)) %>%
    group_by(hours) %>%
    summarize(prop_time_watched = sum(proportion_video)/length(proportion_video)) %>%
    arrange(desc(prop_time_watched))
```

```
## # A tibble: 24 × 2
##    hours prop_time_watched
##    <int>             <dbl>
##  1    12             1.00
##  2    15             0.871
##  3     1             0.829
##  4    19             0.804
##  5    14             0.796
##  6    20             0.778
##  7     6             0.776
##  8    22             0.775
##  9    10             0.765
## 10     9             0.762
## # ℹ 14 more rows
```

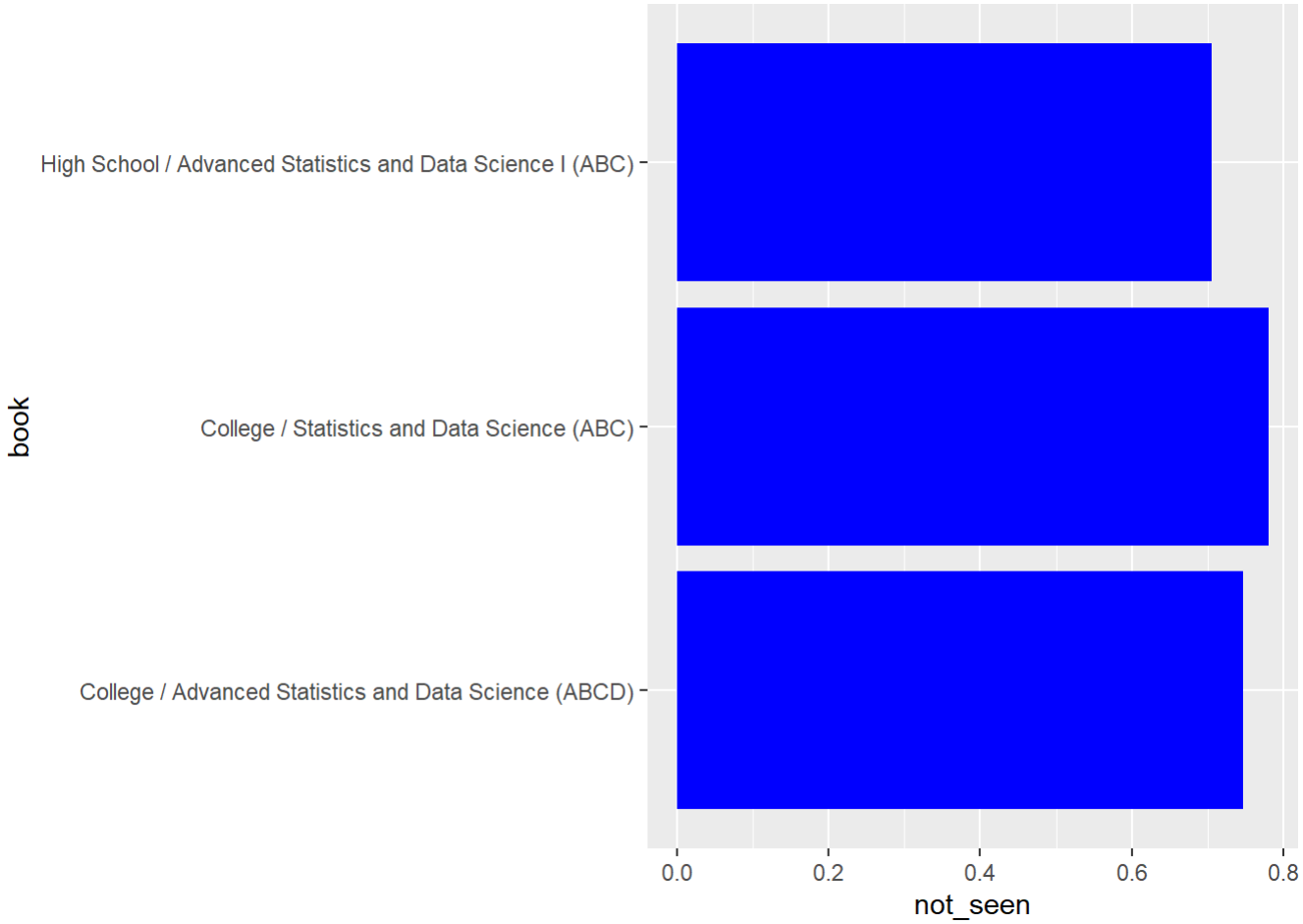
```
media_views %>%
    group_by(chapter_number) %>%
    summarize(not_seen = sum(is.na(proportion_video))/length(proportion_video))
```

```
## # A tibble: 2 × 2
##   chapter_number not_seen
##            <dbl>    <dbl>
## 1              4    0.696
## 2              5    0.812
```

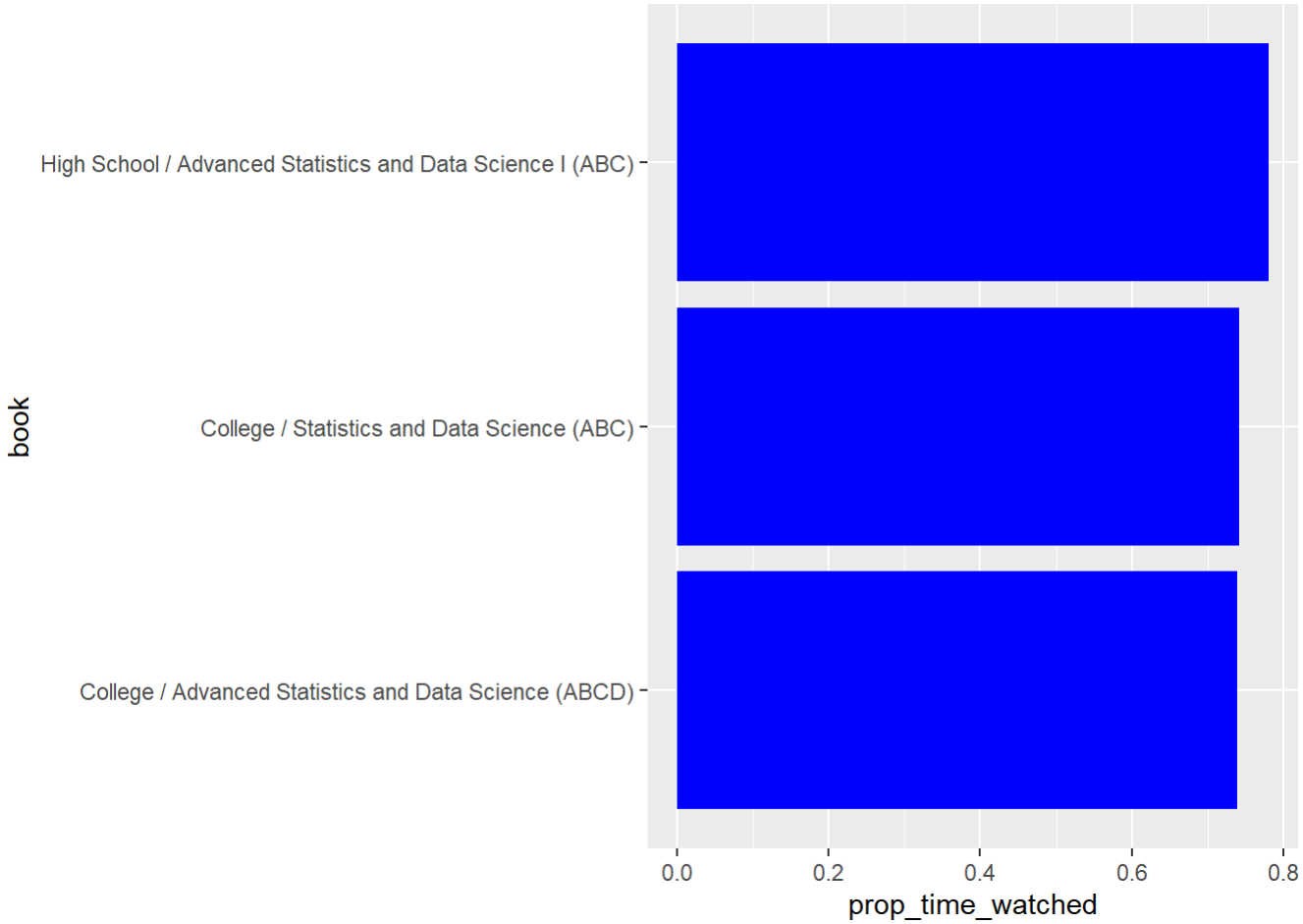
```
media_views %>%
    na.omit() %>%
    group_by(chapter_number) %>%
    summarize(prop_time_watched = sum(proportion_video)/length(proportion_video))
```

```
## # A tibble: 2 × 2
##   chapter_number prop_time_watched
##            <dbl>             <dbl>
## 1              4             0.761
## 2              5             0.721
```

```
#Which book had the least proportion of watches?

#more people are watching these videos in high school compared to college and for longer proport
ions
media_views %>%
    group_by(book) %>%
    summarize(not_seen = sum(is.na(proportion_video))/length(proportion_video)) %>%
    ggplot(aes(book, not_seen)) +
    geom_col(fill = "blue") +
    coord_flip()
```




```
media_views %>%
    na.omit() %>%
    group_by(book) %>%
    summarize(prop_time_watched = sum(proportion_video)/length(proportion_video)) %>%
    ggplot(aes(book, prop_time_watched)) +
    geom_col(fill = "blue") +
    coord_flip()
```




```
#which book and version is viewed the most
media_views %>%
    group_by(book, release) %>%
    summarize(not_seen = sum(is.na(proportion_video))/length(proportion_video)) %>%
    arrange(not_seen)
```

```
## `summarise()` has grouped output by 'book'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 8 × 3
## # Groups:   book [3]
##   book                                                  release   not_seen
##   <chr>                                                 <chr>        <dbl>
## 1 High School / Advanced Statistics and Data Science I (ABC) v5.0         0.706
## 2 College / Advanced Statistics and Data Science (ABCD)      v5.0-exp2    0.738
## 3 College / Statistics and Data Science (ABC)                v5.0-exp1    0.742
## 4 College / Advanced Statistics and Data Science (ABCD)      v5.0         0.743
## 5 College / Statistics and Data Science (ABC)                v5.2         0.763
## 6 College / Advanced Statistics and Data Science (ABCD)      v5.1.1       0.776
## 7 College / Statistics and Data Science (ABC)                v5.0         0.802
## 8 College / Statistics and Data Science (ABC)                v5.1.1       0.947
```

```
#High school textbook has the worst success in EOC questions
checkpoints_eoc %>%
    group_by(book) %>%
    summarize(correctness = mean(EOC, na.rm = TRUE))
```

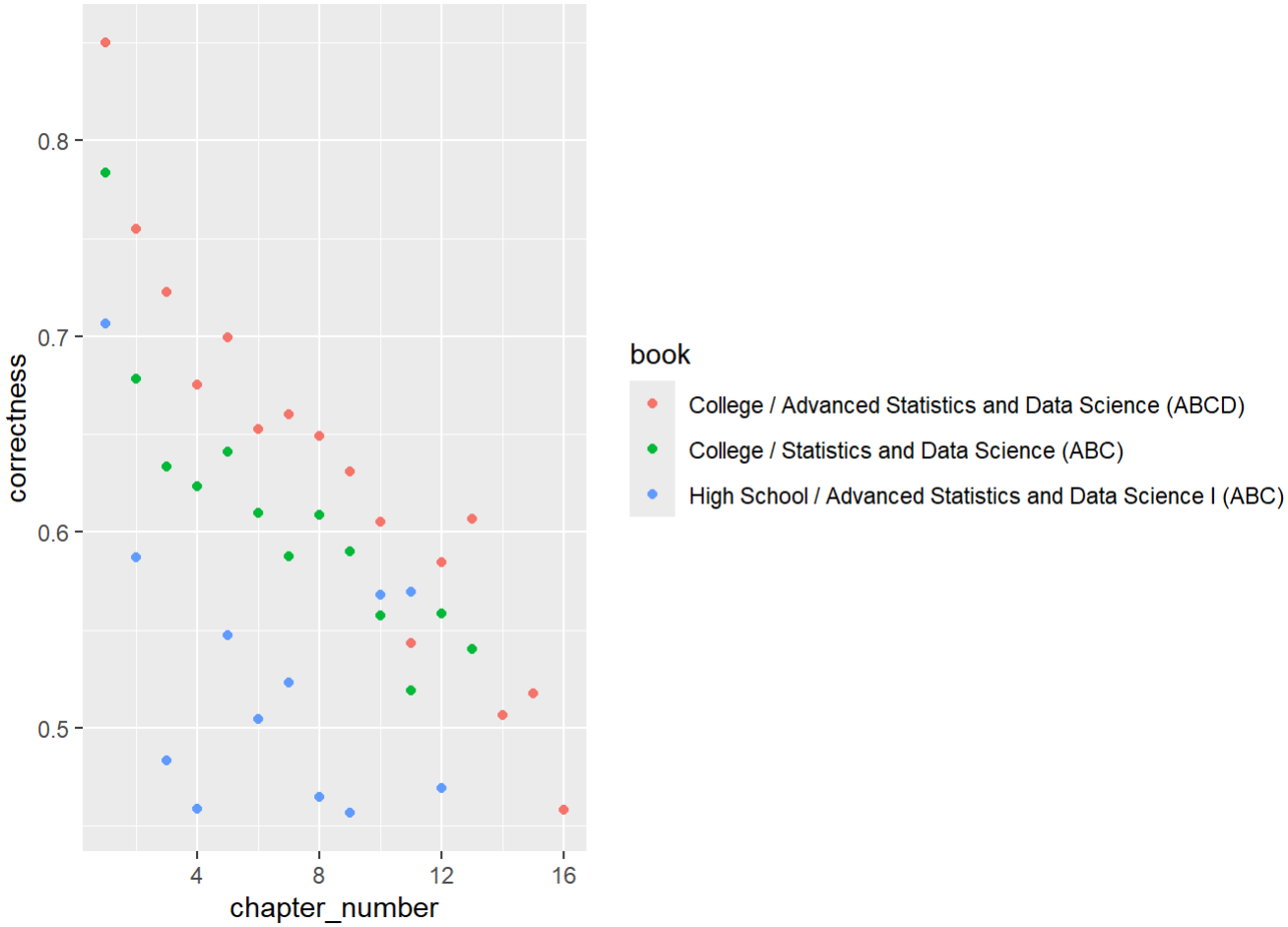
```
## # A tibble: 3 × 2
##   book                                                      correctness
##   <chr>                                                           <dbl>
## 1 College / Advanced Statistics and Data Science (ABCD)           0.657
## 2 College / Statistics and Data Science (ABC)                     0.619
## 3 High School / Advanced Statistics and Data Science I (ABC)      0.527
```

```
#High school textbook is attempting the most times for EOC questions
checkpoints_eoc %>%
    group_by(book) %>%
    summarise(prop_attempts = sum(n_attempt)/sum(n_possible))
```

```
## # A tibble: 3 × 2
##   book                                                      prop_attempts
##   <chr>                                                             <dbl>
## 1 College / Advanced Statistics and Data Science (ABCD)              1.47
## 2 College / Statistics and Data Science (ABC)                        1.55
## 3 High School / Advanced Statistics and Data Science I (ABC)         1.88
```

```
#Students accuracy in answer decreases as you get deeper into the book
checkpoints_eoc %>%
    group_by(chapter_number, book) %>%
    summarize(correctness = mean(EOC, na.rm = TRUE)) %>%
    arrange(correctness) %>%
    ggplot(aes(chapter_number, correctness)) +
    geom_point(aes(color = book))
```

```
## `summarise()` has grouped output by 'chapter_number'. You can override using
## the `.groups` argument.
```




```
checkpoints_eoc %>%
    group_by(class_id, book) %>%
    summarize(correctness = mean(EOC, na.rm = TRUE)) %>%
    arrange(desc(correctness))
```

```
## `summarise()` has grouped output by 'class_id'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 48 × 3
## # Groups:   class_id [48]
##    class_id                             book                          correctness
##    <chr>                                <chr>                               <dbl>
##  1 d1175d28-51bb-44af-b4e0-6f7a36c8cc43 College / Advanced Statisti…       0.772
##  2 1020418a-3eeb-4251-88f7-150c8fe00a56 College / Statistics and Da…       0.745
##  3 51711479-441b-4c02-aef7-517aca63a53f College / Statistics and Da…       0.724
##  4 34036ef8-17eb-4e96-b97e-8ec8054a290f College / Advanced Statisti…       0.713
##  5 074123e7-cd90-4500-86fe-286aaa733bf5 College / Statistics and Da…       0.710
##  6 20bd524c-bb2d-4b74-a419-929475b91d94 College / Statistics and Da…       0.705
##  7 6fbf5a0a-cf5d-4567-89b5-eb5c4a16c4ab College / Advanced Statisti…       0.704
##  8 94da41a4-f9f8-4225-bf41-42db737850b9 College / Statistics and Da…       0.697
##  9 103f5ce8-9e95-4916-815e-9f821d274a59 College / Statistics and Da…       0.697
## 10 c09145c1-d635-41ae-b881-17ab46895fe4 College / Statistics and Da…       0.693
## # ℹ 38 more rows
```

```
checkpoints_eoc %>%
    filter(class_id == "d1175d28-51bb-44af-b4e0-6f7a36c8cc43" | class_id == "1020418a-3eeb-4251-88f7-150c8fe00a56" | class_id == "51711479-441b-4c02-aef7-517aca63a53f") %>%
    distinct(book)
```

```
## # A tibble: 2 × 1
##   book
##   <chr>
## 1 College / Statistics and Data Science (ABC)
## 2 College / Advanced Statistics and Data Science (ABCD)
```

```
checkpoints_eoc %>%
    filter(class_id == "1cca9f91-5c4a-4e1a-8e0e-293b070dfd6f" | class_id == "9bdf8bfc-9998-4fd8-85d2-70c91cf94891" | class_id == "52619962-72f6-4716-9c64-1c06fe10f739") %>%
    distinct(book)
```

```
## # A tibble: 1 × 1
##   book
##   <chr>
## 1 College / Statistics and Data Science (ABC)
```

```
responses %>%
    distinct(book)
```

```
## # A tibble: 1 × 1
##   book
##   <chr>
## 1 College / Advanced Statistics and Data Science (ABCD)
```

```
responses %>%
    count(lrn_type)
```

```
## # A tibble: 9 × 2
##   lrn_type                  n
##   <chr>                 <int>
## 1 association            3726
## 2 choicematrix           6624
## 3 clozeassociation       1030
## 4 imageclozeassociation   408
## 5 mcq                  109275
## 6 plaintext             29564
## 7 shorttext              5802
## 8 sortlist                208
## 9 <NA>                  51592
```

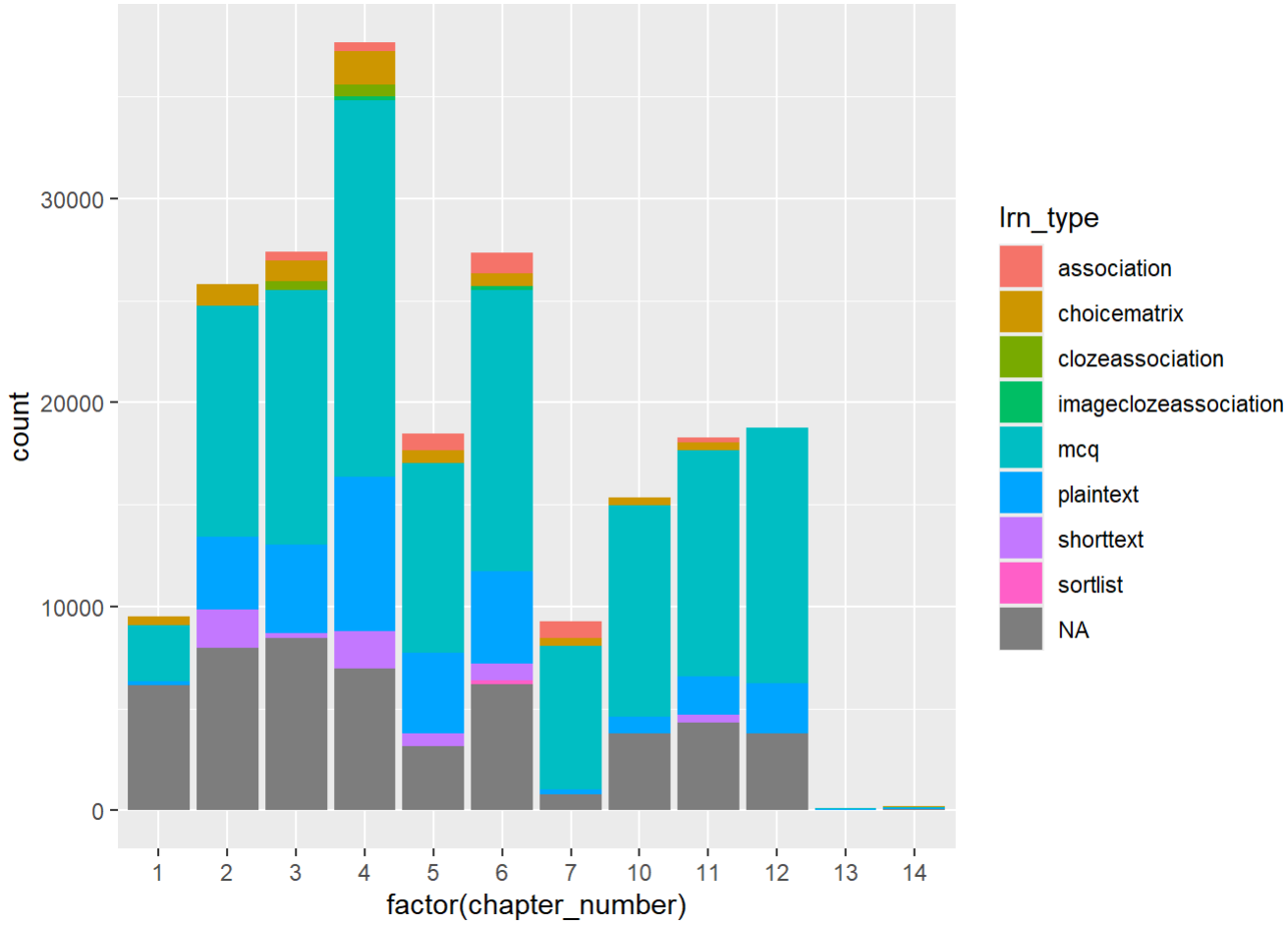
```
responses %>%
    group_by(review_flag) %>%
    summarize(grade = sum(points_earned, na.rm = TRUE)/sum(points_possible, na.rm = TRUE))
```

```
## # A tibble: 2 × 2
##   review_flag grade
##   <lgl>       <dbl>
## 1 FALSE       0.679
## 2 TRUE        0.769
```

```
responses %>%
    group_by(lrn_type) %>%
    summarize(grade = sum(points_earned, na.rm = TRUE)/sum(points_possible, na.rm = TRUE)) %>%
    arrange(desc(grade))
```

```
## # A tibble: 9 × 2
##   lrn_type              grade
##   <chr>                 <dbl>
## 1 association           0.817
## 2 clozeassociation      0.744
## 3 mcq                   0.706
## 4 <NA>                  0.688
## 5 sortlist              0.644
## 6 shorttext             0.551
## 7 imageclozeassociation 0.520
## 8 plaintext             0.504
## 9 choicematrix          0.408
```

```
ggplot(responses, aes(factor(chapter_number), fill = lrn_type)) +
    geom_bar()
```




```
responses %>%
    filter(item_type == "code" & response == prompt)
```

```
## # A tibble: 5,315 × 40
##    book          release chapter page  chapter_number section_number review_flag
##    <chr>         <chr>   <chr>   <chr>          <dbl>          <dbl> <lgl>
##  1 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  2 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  3 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  4 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  5 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  6 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  7 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  8 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  9 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
## 10 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
## # ℹ 5,305 more rows
## # ℹ 33 more variables: institution_id <chr>, class_id <chr>, student_id <chr>,
## #   item_id <chr>, item_type <chr>, response <chr>, prompt <chr>,
## #   points_possible <dbl>, points_earned <dbl>, dt_submitted <dttm>,
## #   completes_page <lgl>, attempt <dbl>, user_agent <chr>,
## #   lrn_response_id <chr>, lrn_activity_reference <chr>,
## #   lrn_question_reference <chr>, lrn_question_position <dbl>, …
```

```
responses %>%
    filter(item_type == "code")
```

```
## # A tibble: 51,592 × 40
##    book          release chapter page  chapter_number section_number review_flag
##    <chr>         <chr>   <chr>   <chr>          <dbl>          <dbl> <lgl>
##  1 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  2 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  3 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  4 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  5 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  6 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  7 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  8 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
##  9 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
## 10 College / Ad… v5.0-e… Chapte… 1.3 …              1              3 FALSE
## # ℹ 51,582 more rows
## # ℹ 33 more variables: institution_id <chr>, class_id <chr>, student_id <chr>,
## #   item_id <chr>, item_type <chr>, response <chr>, prompt <chr>,
## #   points_possible <dbl>, points_earned <dbl>, dt_submitted <dttm>,
## #   completes_page <lgl>, attempt <dbl>, user_agent <chr>,
## #   lrn_response_id <chr>, lrn_activity_reference <chr>,
## #   lrn_question_reference <chr>, lrn_question_position <dbl>, …
```

```
responses %>%
    filter(chapter_number == 9)
```

```
## # A tibble: 0 × 40
## # ℹ 40 variables: book <chr>, release <chr>, chapter <chr>, page <chr>,
## #   chapter_number <dbl>, section_number <dbl>, review_flag <lgl>,
## #   institution_id <chr>, class_id <chr>, student_id <chr>, item_id <chr>,
## #   item_type <chr>, response <chr>, prompt <chr>, points_possible <dbl>,
## #   points_earned <dbl>, dt_submitted <dttm>, completes_page <lgl>,
## #   attempt <dbl>, user_agent <chr>, lrn_response_id <chr>,
## #   lrn_activity_reference <chr>, lrn_question_reference <chr>, …
```

```
page_views %>%
    group_by(page) %>%
    summarize(avg_off_page = mean(off_page_long, na.rm = TRUE)) %>%
    mutate(avg_off_page = avg_off_page / 1000 / 60) %>%
    arrange(desc(avg_off_page))
```

```
## # A tibble: 171 × 2
##    page                                                  avg_off_page
##    <chr>                                                        <dbl>
##  1 Practice Exam Page 2                                          47.7
##  2 14.1 Targeted Model Comparisons                               47.3
##  3 12.1 From Hypothesis Testing to Confidence Intervals          46.6
##  4 Midterm 1 Page 2                                              42.5
##  5 Practice Exam Page 1                                          42.0
##  6 11.8 Pairwise Comparisons                                     38.4
##  7 8.1 Extending to a Three-Group Model                          37.7
##  8 8.8 Chapter 8 Review Questions                                36.5
##  9 15.1 Dogs in the Emergency Room                               36.1
## 10 9.1 Using a Quantitative Explanatory Variable in a Model      35.9
## # ℹ 161 more rows
```

```
page_views %>%
    distinct(book)
```

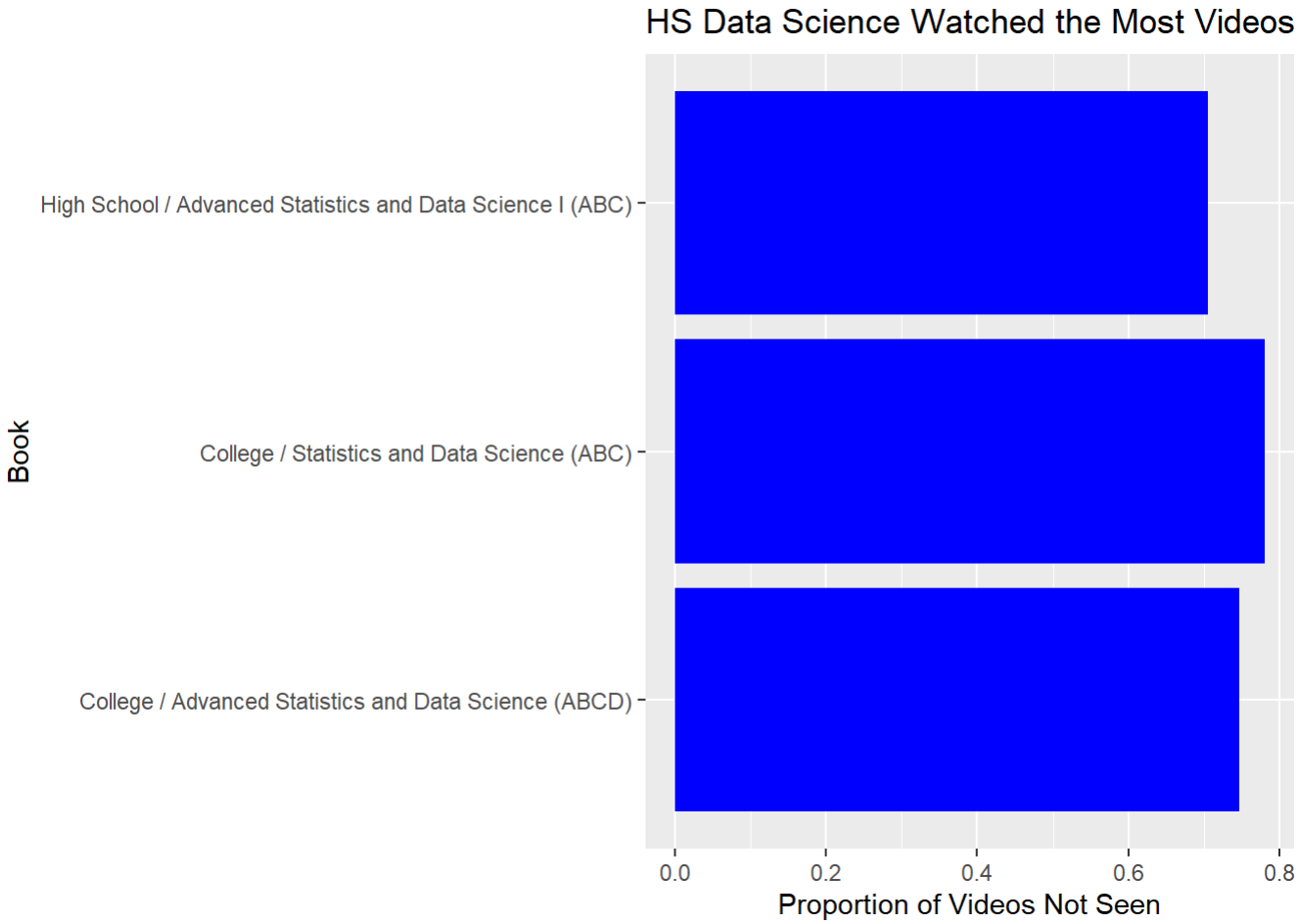
```
## # A tibble: 3 × 1
##   book
##   <chr>
## 1 College / Statistics and Data Science (ABC)
## 2 College / Advanced Statistics and Data Science (ABCD)
## 3 High School / Advanced Statistics and Data Science I (ABC)
```

```
page_views %>%
    group_by(book, release) %>%
    summarize(avg_engaged = mean(engaged, na.rm = TRUE)) %>%
    arrange(desc(avg_engaged)) %>%
    mutate(avg_engaged = avg_engaged / 1000)
```

```
## `summarise()` has grouped output by 'book'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 8 × 3
## # Groups:   book [3]
##   book                                                  release avg_engaged
##   <chr>                                                 <chr>         <dbl>
## 1 College / Advanced Statistics and Data Science (ABCD) v5.0-e…        347.
## 2 College / Statistics and Data Science (ABC)           v5.0-e…        343.
## 3 High School / Advanced Statistics and Data Science I (ABC) v5.0      317.
## 4 College / Statistics and Data Science (ABC)           v5.2           310.
## 5 College / Statistics and Data Science (ABC)           v5.0           303.
## 6 College / Advanced Statistics and Data Science (ABCD) v5.0           298.
## 7 College / Advanced Statistics and Data Science (ABCD) v5.1.1         290.
## 8 College / Statistics and Data Science (ABC)           v5.1.1         280.
```

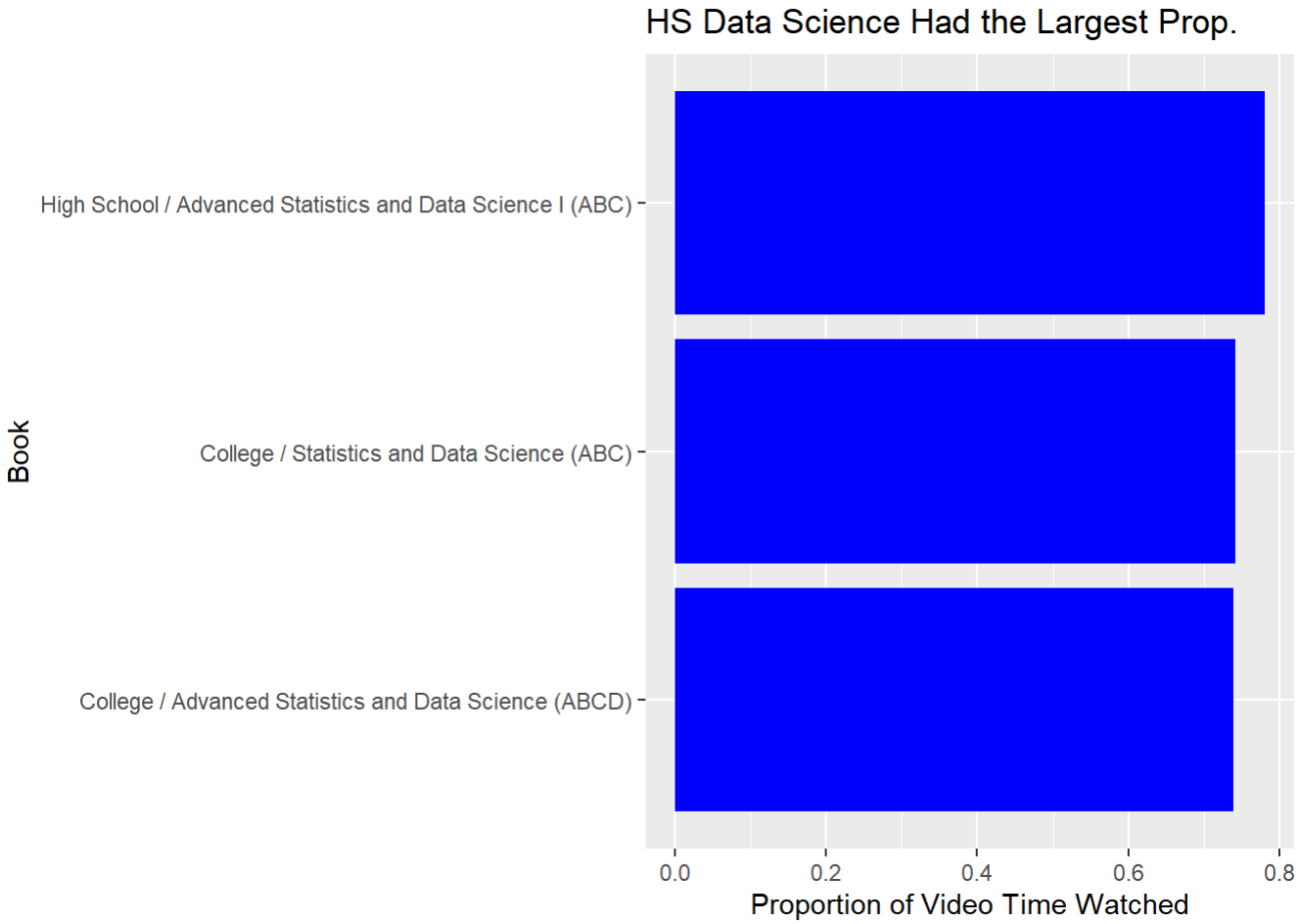
```
#Which book had the least proportion of watches?

#more people are watching these videos in high school compared to college and for longer proport
ions
media_views %>%
    group_by(book) %>%
    summarize(not_seen = sum(is.na(proportion_video))/length(proportion_video)) %>%
    ggplot(aes(book, not_seen)) +
    geom_col(fill = "blue") +
    coord_flip() +
    labs(
        x = "Book",
        y = "Proportion of Videos Not Seen",
        title = "HS Data Science Watched the Most Videos"
    )
```




```
ggsave("Prop_videos.pdf")
```

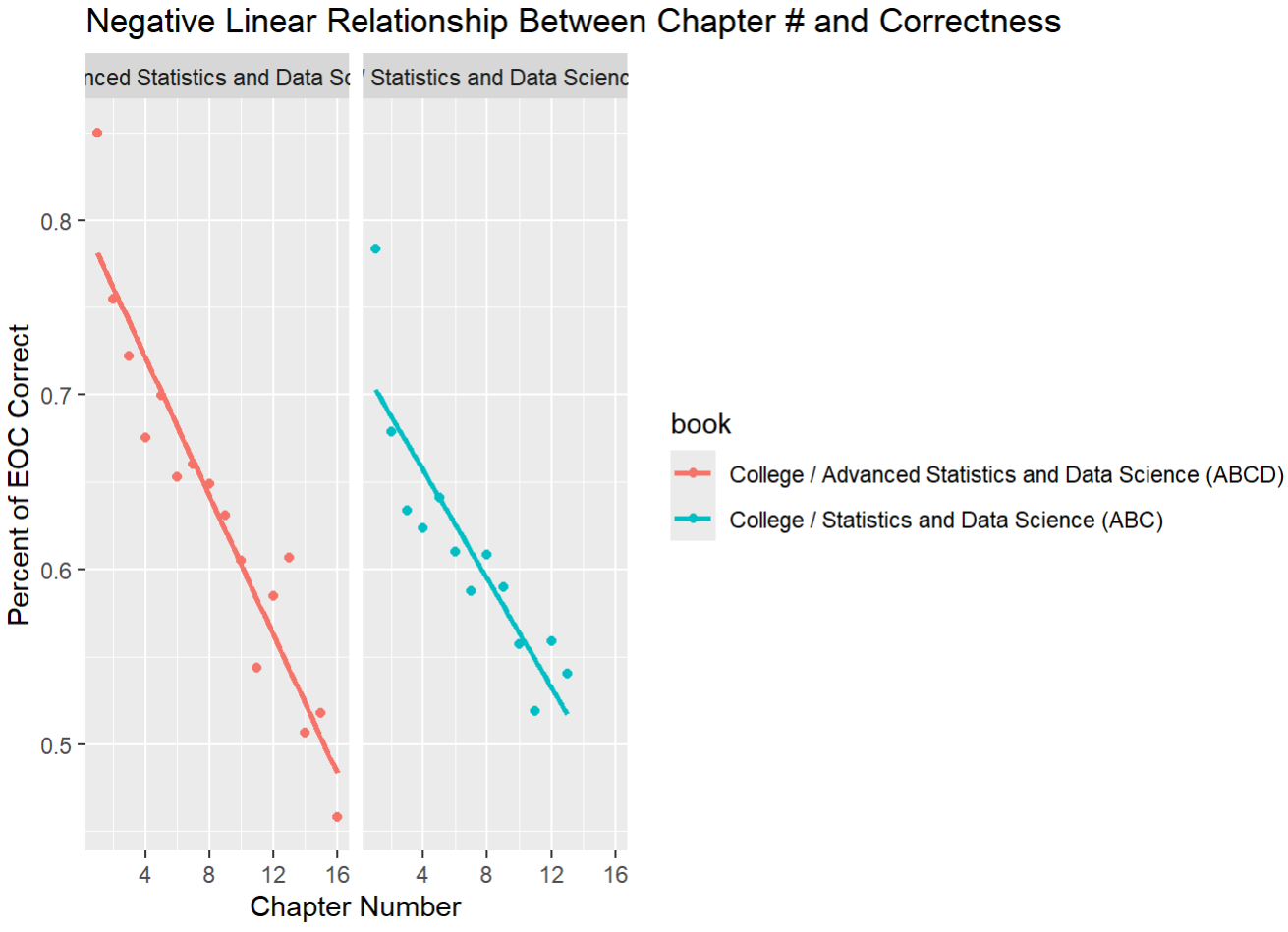
```
## Saving 7 x 5 in image
```

```
media_views %>%
    na.omit() %>%
    group_by(book) %>%
    summarize(prop_time_watched = sum(proportion_video)/length(proportion_video)) %>%
    ggplot(aes(book, prop_time_watched)) +
    geom_col(fill = "blue") +
    coord_flip()+
    labs(
        x = "Book",
        y = "Proportion of Video Time Watched",
        title = "HS Data Science Had the Largest Prop."
    )
```



```
#Students accuracy in answer decreases as you get deeper into the book
checkpoints_eoc %>%
    filter(book != "High School / Advanced Statistics and Data Science I (ABC)") %>%
    group_by(chapter_number, book) %>%
    summarize(correctness = mean(EOC, na.rm = TRUE)) %>%
    arrange(correctness) %>%
    ggplot(aes(chapter_number, correctness, group = book)) +
    geom_point(aes(color = book)) +
    facet_grid(~ book) +
    geom_smooth(method = lm, aes(color = book), se = FALSE) +
    labs(
        x = "Chapter Number",
        y = "Percent of EOC Correct",
        title = "Negative Linear Relationship Between Chapter # and Correctness"
    )
```

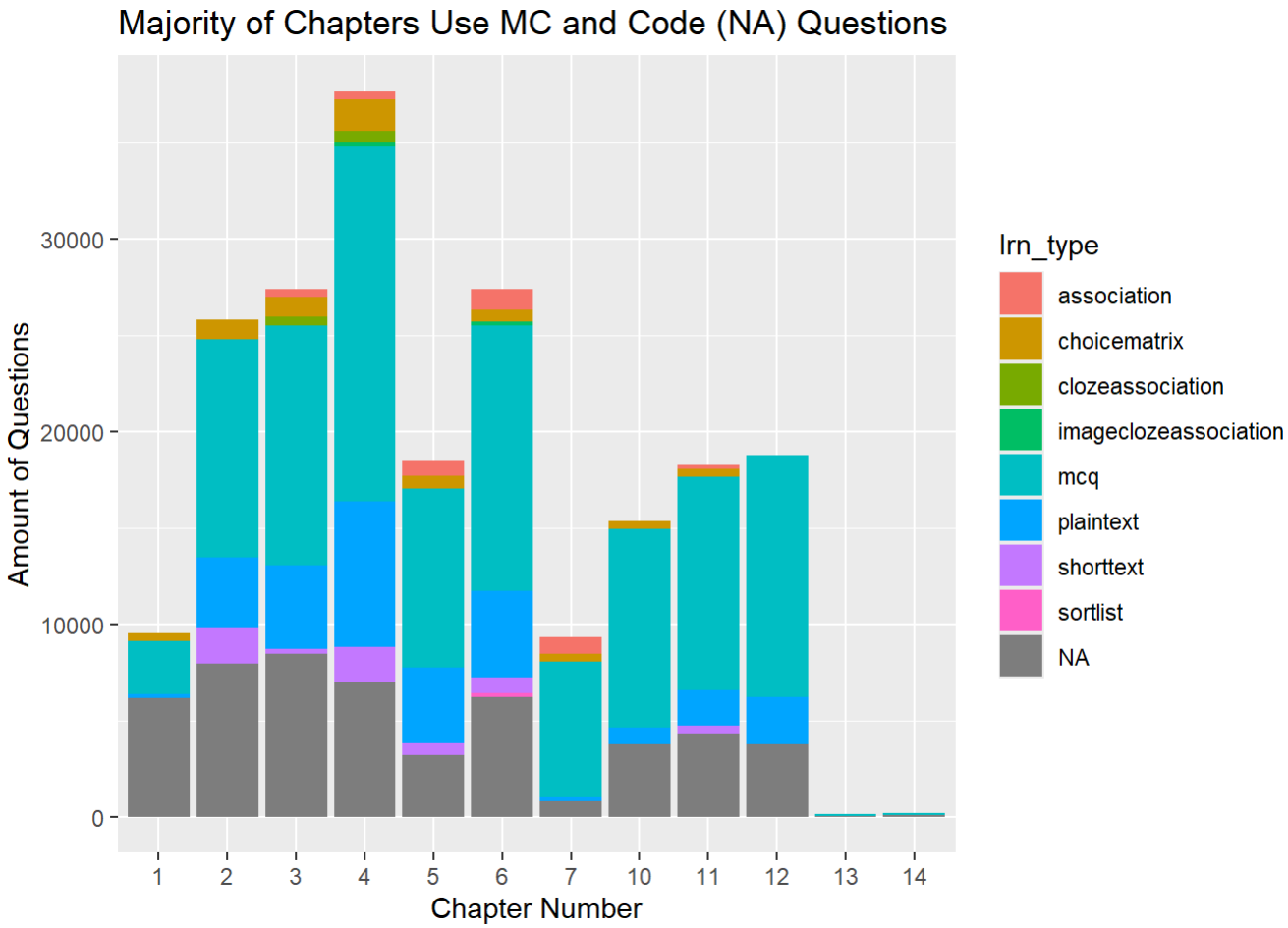
```
## `summarise()` has grouped output by 'chapter_number'. You can override using
## the `.groups` argument.
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggsave("eoc.pdf")
```

```
## Saving 7 x 5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

```
ggplot(responses, aes(factor(chapter_number), fill = lrn_type)) +
    geom_bar() +
    labs(
        x = "Chapter Number",
        y = "Amount of Questions",
        title = "Majority of Chapters Use MC and Code (NA) Questions"
    )
```




```
ggsave("Learntype.pdf")
```

```
## Saving 7 x 5 in image
```

```
checkpoints_pulse %>%
    filter(construct == "Utility Value") %>%
    na.omit() %>%
    mutate(chapter_number = chapter_number - 1,
           disagree = if_else(response == 0 | response == 1 | response == 2, TRUE, FALSE)) %>%
    group_by(chapter_number) %>%
    summarize(
        prop_disagreed = mean(disagree) * 100
    ) %>%
    arrange(desc(prop_disagreed))
```

```
## # A tibble: 12 × 2
##    chapter_number prop_disagreed
##             <dbl>          <dbl>
##  1              8           5.46
##  2              7           5.09
##  3              2           4.88
##  4              3           4.60
##  5              6           4.39
##  6             10           4.38
##  7             11           3.98
##  8              5           3.93
##  9              9           3.90
## 10              4           3.79
## 11              1           1.88
## 12             12           1.57
```

```
checkpoints_pulse %>%
    distinct(construct)
```

```
## # A tibble: 4 × 1
##   construct
##   <chr>
## 1 Cost
## 2 Expectancy
## 3 Intrinsic Value
## 4 Utility Value
```

```
checkpoints_pulse %>%
    filter(construct == "Expectancy") %>%
    na.omit() %>%
    mutate(chapter_number = chapter_number - 1,
           disagree = if_else(response == 0 | response == 1 | response == 2, TRUE, FALSE)) %>%
    group_by(chapter_number) %>%
    summarize(
        prop_disagreed = mean(disagree) * 100
    ) %>%
    arrange(desc(prop_disagreed))
```

```
## # A tibble: 12 × 2
##    chapter_number prop_disagreed
##             <dbl>          <dbl>
##  1              8           9.59
##  2              2           9.38
##  3              7           9.33
##  4             10           8.42
##  5             11           8.38
##  6              3           8.12
##  7              6           7.88
##  8              9           7.06
##  9             12           6.67
## 10              5           6.63
## 11              4           6.51
## 12              1           3.16
```

```
checkpoints_pulse %>%
    filter(construct == "Cost") %>%
    na.omit() %>%
    mutate(chapter_number = chapter_number - 1,
           disagree = if_else(response == 0 | response == 1 | response == 2, TRUE, FALSE)) %>%
    group_by(chapter_number) %>%
    summarize(
        prop_disagreed = mean(disagree) * 100
    ) %>%
    arrange(desc(prop_disagreed))
```

```
## # A tibble: 12 × 2
##    chapter_number prop_disagreed
##             <dbl>          <dbl>
##  1              1           52.4
##  2             12           47.8
##  3             10           41.2
##  4              2           40.4
##  5              3           40.0
##  6              9           39.2
##  7             11           37.7
##  8              8           37.4
##  9              6           36.9
## 10              5           36.7
## 11              7           36.2
## 12              4           35.4
```

```
checkpoints_pulse %>%
    group_by(chapter_number) %>%
    summarize(
        did_not_respond = sum(is.na(response))/length(response)
    ) %>%
    arrange(desc(did_not_respond)) %>%
    ggplot(aes(factor(chapter_number), did_not_respond)) +
    geom_col() +
    labs(
        x = "Chapter Number",
        y = "Proportion Who Did Not Respond",
        title = "Later Chapters Did Not Respond to Pulse Questions as Much as Earlier Chapters"
    )
```