CDPH Course Analytics

David Connell and Michelle Ruiz

2023-11-06

R/RStudio Fundamentals, on-demand ongoing offering

Enrollment

1. Download entire enrollment CSV from catalog analytics
2. Rename CSV file "enrollment_export.csv"
3. Move CSV file into course-analytics folder

463 total participants enrolled in R/RStudio Fundamentals as of 2023-11-06

4 participants enrolled in R/RStudio Fundamentals since 2023-10-30

Certificates

TODO: add instructions for getting CSV from certificate generator website

1. Go to certifier.io, go to "Credentials" and select all, export
2. Rename CSV file "certificates.csv"
3. Move CSV file into course-analytics folder

48 received a certificate of completion for R/RStudio Fundamentals as of 2023-11-06 and **5** received it since 2023-10-30

Viewership

list of videos and how many have seen each video total

how many have seen all 5 seen 4 seen 3 seen less than 3

Data Visualization, August-December 2024 offering

1. Download **entire gradebook** from canvas as CSV
2. Rename CSV file to remove date
3. Move CSV file into course-analytics folder

Enrollment

387 total participants enrolled in Data Visualization as of 2023-11-06

3 participants enrolled in Data Visualization since 2023-10-30

Viewership

```
viz_grades <- read_csv("Grades-Data_Visualization_for_Public_Health.csv")
```

```
## Rows: 378 Columns: 43
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (21): Student, SIS Login ID, Section, Course Videos Current Points, Cour...
## dbl (21): ID, Introduction (1647), Beyond the Bar Graph (1643), Introduction...
## lgl  (1): SIS User ID
##
## 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.
```

```
viz_video_names <- names(viz_grades[seq(from = 6, to = 24)])
viz_video_views <- tibble(video_names = viz_video_names,
                          views = NA)


dim(filter(viz_grades, `Introduction (1647)` == 1))[1]
```

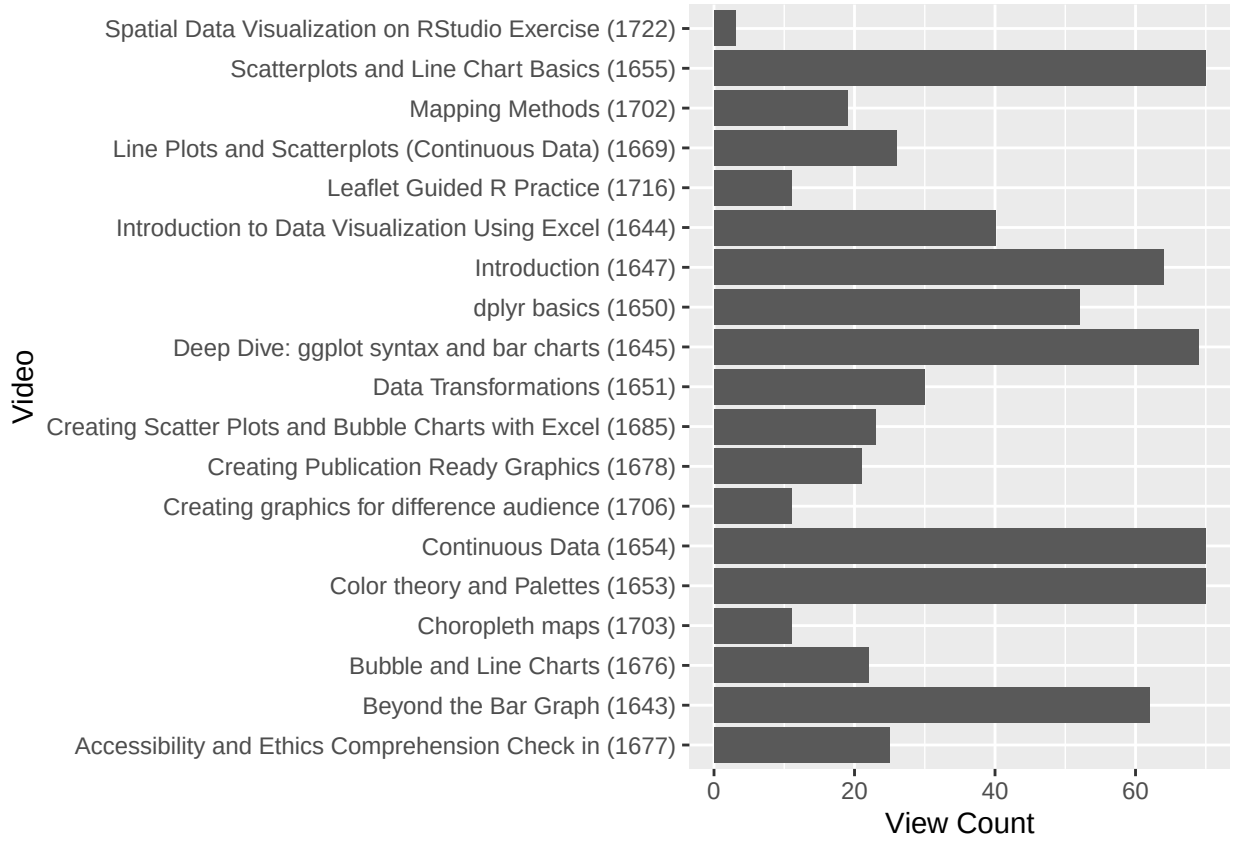
```
## [1] 64
```

```
temp <- viz_grades %>%
    group_by(across(6:24)) %>%
    summarise(n = n())
```

```
## `summarise()` has grouped output by 'Introduction (1647)', 'Beyond the Bar
## Graph (1643)', 'Introduction to Data Visualization Using Excel (1644)', 'Deep
## Dive: ggplot syntax and bar charts (1645)', 'Creating Scatter Plots and Bubble
## Charts with Excel (1685)', 'dplyr basics (1650)', 'Data Transformations
## (1651)', 'Color theory and Palettes (1653)', 'Continuous Data (1654)',
## 'Scatterplots and Line Chart Basics (1655)', 'Line Plots and Scatterplots
## (Continuous Data) (1669)', 'Bubble and Line Charts (1676)', 'Accessibility and
## Ethics Comprehension Check in (1677)', 'Creating Publication Ready Graphics
## (1678)', 'Mapping Methods (1702)', 'Choropleth maps (1703)', 'Creating graphics
## for difference audience (1706)', 'Leaflet Guided R Practice (1716)'. You can
## override using the `.groups` argument.
```

```
# Assuming your data frame is named df
# First, let's convert the data frame from wide to long format
df_long <- viz_grades %>%
    select(-(2:5)) %>%
    select(-(21:39)) %>%
```

```r
  pivot_longer(
    cols = 2:20,
    names_to = "Video",  # This will be the new column with video names
    values_to = "Watched" # This will be the new column with the values 1 or 0
  ) %>%
    filter(Watched == 1) # Keep only rows where the video was watched




# Now, we can summarize the data to count viewership
video_count <- df_long %>%
  count(Video) %>%
  arrange(desc(n)) # Optional: Arrange in descending order of counts

video_count_new <- video_count %>%
    mutate(order = case_when(
        Video == "Introduction (1647)" ~ 1,
        Video == "Color theory and Palettes (1653)" ~ 3)) %>%
    arrange(order)



# Now let's create a bar chart with ggplot2
ggplot(video_count, aes(y = Video, x = n)) +
  geom_bar(stat = "identity") +
  labs(y = "Video", x = "View Count")
```



How many people have watched each of the videos in the course?

Since there are many videos for this course, this code be summarized in a graphic where on the x axis you have video number (organized chronologically) and the y axis is the completion rate or the number of people who have completed the video. If there are many videos each week some indicator of the corresponding week can be added to the plot. o When the plot is made, denote the current date (when the data was pulled) and add a dashed vertical line to the plot in between videos that occur before and after that date.

Causal Inference, July-September 2024 offering

Enrollment

815 total participants enrolled in Introduction to Causal Inference for Public Health Professionals as of 2023-11-06

3 participants enrolled in Introduction to Causal Inference for Public Health Professionals since 2023-10-30

Viewership