STAT 1261 Individual Project

Clare Cruz

10/14/2020

Recreation of the Tarantino Graph:

Article Title: "A Complete Catalog Of Every Time Someone Cursed Or Bled Out In A Quentin Tarantino Movie"

http://fivethirtyeight.com/features/complete-catalog-curses-deaths-quentin-tarantino-films/

Quentin Tarantino movies are known for being full of swear words and lots of blood. This data set quantifies the amount of profanity and death by stating when a character said a swear word and when someone died throughout Tarantino's most popular movies. The data set includes 7 movies, every swear word you could think of, and the minute time in that movie when the death or swear word occurred.

In terms of data wrangling, this data set required little work because it was already tidy. I did have to mutate the profane column from a boolean to a character value so that the legend would come up correctly. I also had to integrate the year of the movie to properly sort and arrange the facets in ascending order according to the year the movie was produced.

While a dot plot might appear to be easy to work with, there are many technical obstacles that comes with the geom function. The biggest obstacle is that the dot plot does not actually give a count of the dots as a y-axis, as one would expect (similar in a histogram chart). This obstacle is a known error on github website, with no indication of an update in the future (https://github.com/tidyverse/ggplot2/issues/2203). Therefore, I had to force the axis to be from 0-20, which ends up correctly displaying the data.

I also had to force the x-axis so that there was no extra area to the left and right of each facet. One of the hardest parts were figuring out how to get the dots to appear correctly and adjusting the heights of the facets so that it was produced correctly. I had to fill and color the dots in the same way, which required two separate functions each time color was mentioned. Additionally, I had to adjust the size and position of the facet labels to match the graph. Next, the legend needed to be placed in an unconventional position, so I had to increase the margins and make room to put the legend between the title and the graphs themselves.

The absolute hardest part was adding annotations. Because of the error mentioned above with no y-axis, it was impossible to use the any traditional geom_text function to add individual annotation. After a long search, the draw_label function in cowplot allowed me to enter texts on top of the entire plot. I also added rectangles to cover the ends of the facets seen in the graphs. The only thing I was unable to figure out was the connecting lines between the annotations and the chart. However, I manipulated text labels to mimic this as best as I could. Because of this, the manipulated arrows appear slightly different in html and pdf forms. Please consider the pdf product of the program.

```
library(fivethirtyeight)
```

```
## Some larger datasets need to be installed separately, like senators and
## house_district_forecast. To install these, we recommend you install the
## fivethirtyeightdata package by running:
## install.packages('fivethirtyeightdata', repos =
## 'https://fivethirtyeightdata.github.io/drat/', type = 'source')
```

```r
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
```

```r
library(grid)
library(tidyr)
library(grid)
library(cowplot)#lifesaver
head(fivethirtyeight::tarantino)
```

```
## # A tibble: 6 x 4
##   movie          profane word     minutes_in
##   <chr>          <lgl>   <chr>         <dbl>
## 1 Reservoir Dogs TRUE    dick           0.4
## 2 Reservoir Dogs TRUE    dicks          0.43
## 3 Reservoir Dogs TRUE    fucked         0.55
## 4 Reservoir Dogs TRUE    fucking        0.61
## 5 Reservoir Dogs TRUE    bullshit       0.61
## 6 Reservoir Dogs TRUE    fuck           0.66
```

```r
#Adding the year information into the movie variables
movie <- c("Reservoir Dogs", "Pulp Fiction", "Jackie Brown", "Kill Bill: Vol. 1", "Kill Bill: Vol. 2",
start_year <- c(1992,1994,1997,2003,2004,2009,2012)
year_info <- data.frame(movie,start_year)
```

```r
tarantino_df <-fivethirtyeight::tarantino %>%
  #Change true false to profanity and death
  mutate(profane, profane = ifelse(profane == T, "PROFANITY", "DEATH"))%>%
  #Adding the year info to the chart
  left_join(year_info)%>%
  #Combine year information with movie
  unite(movie, movie, start_year, sep = ", ")
```

```
## Joining, by = "movie"
```

```r
#turn movie into a factor variable so it can be used in facet wrap
tarantino_df$movie <- factor(tarantino_df$movie, levels = c("Reservoir Dogs, 1992", "Pulp Fiction, 1994

#Re-level levels so that the movies are in all capgs
levels(tarantino_df$movie) <- c("RESERVOIR DOGS, 1992", "PULP FICTION, 1994", "JACKIE BROWN, 1997", "KI
```

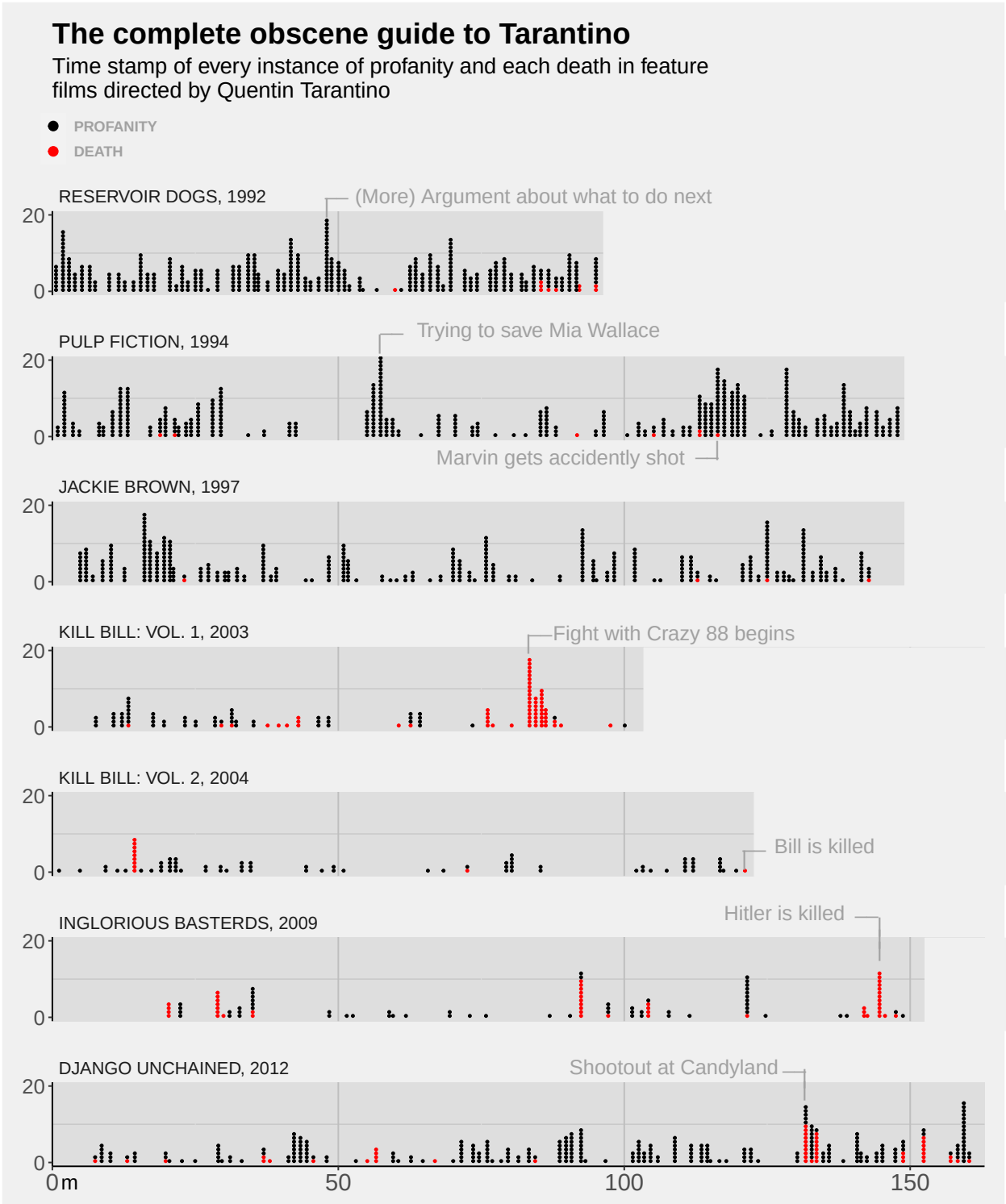
```r
#structure of rectangle used to cut off facets
g <- grid::rectGrob(gp = grid::gpar(fill = "#f0f0f0", col = "#f0f0f0"))
```

```r
tarantino_plot <- ggplot(tarantino_df, aes(x = minutes_in, fill = profane)) +
  geom_dotplot(
    binwidth = 1,
```

```
  method = "dotdensity",
  stackratio = 1.35, #changes spacing between the dots
  aes(color = profane), #how the dots are colored
  dotsize = 0.5,
  stackgroups = TRUE,
  binpositions="all"
  )+
facet_wrap(~movie, ncol = 1, shrink = TRUE
  ) +
#fill and outline of the dots are red and black
scale_color_manual(values =  c("red", "black"))+
scale_fill_manual(values =  c("red", "black")
  )+
#weird x and y axis are formatted to reflect graph
#breaks = what show up on tick marks, limits are max and min
scale_y_continuous(breaks=c(0, 20), limits = c(0, 20))+
scale_x_continuous(expand = c(0, 0),
                   breaks = c(0,50,100,150),
                   limits = c(0, 163)
  )+
#Changed the order of the dots in the legend
guides(
      color = guide_legend(reverse = TRUE),
       fill = guide_legend(reverse = TRUE)
  )+
theme(#position for the legend
      legend.position=c(0.05,1.08),
      #no legend title
      legend.title =element_blank(),
      #increases margins to place legend
      plot.margin = unit(c(0.5,0.5,0.5,0.5),"cm"),
      #increases space between the facets
      panel.spacing = unit(1.5, "lines"),
      #background of labels is the same as background
      strip.background =element_rect(fill='#f0f0f0'),
      #background of legend is the same as background
      legend.background =element_rect(fill='#f0f0f0'),
      #text for the legend is colored, bold, and large
      legend.text = element_text(colour = "#a1a1a1", face = "bold", size = 10),
      #Facet labels are adjusted to the left
      strip.text = element_text(hjust = 0),
      #X axis values are large and bold
      axis.text.x = element_text(size = 16),
      #Y axis values are large and bold
      axis.text.y = element_text(size = 14),
      #Facet labels are large
      strip.text.x = element_text(size = 12),
      #minor x axis lines are colored
      panel.grid.minor.x = element_line(colour = "#dedede"),
      #minor y axis lines are colored
      panel.grid.minor.y = element_line(colour = "grey"),
      #major x axis lines are colored
      panel.grid.major.x = element_line(colour = "grey"),
```

```
        #major y axis lines are colored
        panel.grid.major.y = element_line(colour = "#dedede"),
        #background of facet plots are colored
        panel.background = element_rect(fill = "#dedede"),
        #background of the entire plot is colored
        plot.background = element_rect(fill = '#f0f0f0'),
        #axis line is black
        axis.line = element_line(colour = "black"),
        #title of plot is bolded and large
        plot.title = element_text(size = 20, face = "bold"),
        #subtitle of plot is slightly large
        plot.subtitle = element_text(size = 15)
        )+
        labs(title = "The complete obscene guide to Tarantino",
              subtitle = "Time stamp of every instance of profanity and each death in feature\nfilms di


ggdraw(tarantino_plot) +
  #Annotation
  draw_label("(More) Argument about what to do next", x = 0.53, y = 0.84, color = "#a1a1a1") +
  draw_label("|", x = 0.324, y = 0.832, color = "#a1a1a1")+
  draw_label("__", x = 0.334, y = 0.847, color = "#a1a1a1")+

  #Annotation
  draw_label("Trying to save Mia Wallace", x = 0.535, y = 0.73, color = "#a1a1a1")+
  draw_label("|", x = 0.377, y = 0.72, color = "#a1a1a1")+
  draw_label("__", x = 0.389, y = 0.735, color = "#a1a1a1")+

  #Annotation
  draw_label("Marvin gets accidently shot", x = 0.557, y = 0.625, color = "#a1a1a1") +
  draw_label("|", x = 0.713, y = 0.632, color = "#a1a1a1")+
  draw_label("__", x = 0.703, y = 0.633, color = "#a1a1a1")+

  #Annotation
  draw_label("Fight with Crazy 88 begins", x = 0.67, y = 0.48, color = "#a1a1a1")+
  draw_label("|", x = 0.525, y = 0.473, color = "#a1a1a1")+
  draw_label("__", x = 0.537, y = 0.487, color = "#a1a1a1")+

  #Annotation
  draw_label("Hitler is killed", x = 0.78, y = 0.25, color = "#a1a1a1")+
  draw_label("|", x = 0.875, y = 0.24, color = "#a1a1a1")+
  draw_label("|", x = 0.875, y = 0.23, color = "#a1a1a1")+
  draw_label("|", x = 0.875, y = 0.22, color = "#a1a1a1")+
  draw_label("__", x = 0.862, y = 0.255, color = "#a1a1a1")+

  #Annotation
  draw_label("Shootout at Candyland", x = 0.67, y = 0.123, color = "#a1a1a1") +
  draw_label("|", x = 0.80, y = 0.115, color = "#a1a1a1")+
  draw_label("|", x = 0.80, y = 0.105, color = "#a1a1a1")+
  draw_label("__", x = 0.7899, y = 0.128, color = "#a1a1a1")+

  #Rectangles to cut off graphs
  draw_grob(g, x = 0.6, y = 0.753, height = 0.075, width = 0.4)+
```

```
  draw_grob(g, x = 0.9, y = 0.633, height = 0.075, width = 0.1)+
  draw_grob(g, x = 0.9, y = 0.513, height = 0.075, width = 0.1)+
  draw_grob(g, x = 0.64, y = 0.393, height = 0.075, width = 0.4)+
  draw_grob(g, x = 0.75, y = 0.273, height = 0.076, width = 0.25)+
  draw_grob(g, x = 0.92, y = 0.153, height = 0.076, width = 0.1)+

  #Annotation with text over it
  draw_label("Bill is killed", x = 0.82, y = 0.305, color = "#a1a1a1") +
  draw_label("|", x = 0.74, y = 0.295, color = "#a1a1a1")+
  draw_label("__", x = 0.75, y = 0.31, color = "#a1a1a1")+

  draw_label("m", x = 0.068, y = 0.028, color = "black")
```


The complete obscene guide to Tarantino
Time stamp of every instance of profanity and each death in feature films directed by Quentin Tarantino
PROFANITY
DEATH
RESERVOIR DOGS, 1992
(More) Argument about what to do next
PULP FICTION, 1994
Trying to save Mia Wallace
Marvin gets accidently shot
JACKIE BROWN, 1997
KILL BILL: VOL. 1, 2003
Fight with Crazy 88 begins
KILL BILL: VOL. 2, 2004
Bill is killed
INGLORIOUS BASTERDS, 2009
Hitler is killed
DJANGO UNCHAINED, 2012
Shootout at Candyland
20
0
0m
50
100
150