

Exploratory Analysis

Mariachiara Fortuna

March 16, 2018

R basics

```
vec <- c(1, 2, 6)

vec

## [1] 1 2 6

char_vec <- c("a", "b", "a", "a", "b")

fac_vec <- as.factor(char_vec)
```

Data import

```
file_name <- file.path("data", "dc-wikia-data.csv")

dc <- read.csv(file_name)

dc <- read.csv("data/dc-wikia-data.csv", na.string = "")
```

Data exploration - Basics

```
View(dc)

str(dc)
```

```
## 'data.frame': 6896 obs. of 13 variables:
## $ page_id : int 1422 23387 1458 1659 1576 1448 1486 1451 71760 1380 ...
## $ name : Factor w/ 6896 levels "3g4 (New Earth)",...: 593 6007 2487 2996 5278 6772 378 62...
## $ urlslug : Factor w/ 6896 levels "\\wiki\\3g4_(New_Earth)",...: 597 6006 2488 2997 5277 6...
## $ ID : Factor w/ 3 levels "Identity Unknown",...: 3 3 3 2 3 2 2 3 2 3 ...
## $ ALIGN : Factor w/ 4 levels "Bad Characters",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ EYE : Factor w/ 17 levels "Amber Eyes","Auburn Hair",...: 4 4 5 5 4 4 4 4 4 4 ...
## $ HAIR : Factor w/ 17 levels "Black Hair","Blond Hair",...: 1 1 4 17 1 1 2 1 2 2 ...
## $ SEX : Factor w/ 4 levels "Female Characters",...: 3 3 3 3 3 1 3 3 1 3 ...
## $ GSM : Factor w/ 2 levels "Bisexual Characters",...: NA NA NA NA NA NA NA NA NA NA ...
## $ ALIVE : Factor w/ 2 levels "Deceased Characters",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ APPEARANCES : int 3093 2496 1565 1316 1237 1231 1121 1095 1075 1028 ...
## $ FIRST.APPEARANCE: Factor w/ 774 levels "1935, October",...: 15 455 156 461 20 33 39 486 261 129 ..
## $ YEAR : int 1939 1986 1959 1987 1940 1941 1941 1989 1969 1956 ...
```

Statistiche descrittive

Var quantitative

```
min(dc$YEAR, na.rm = T)

## [1] 1935

max(dc$YEAR, na.rm = T)

## [1] 2013

range(dc$YEAR, na.rm = T)

## [1] 1935 2013

mean(dc$APPEARANCES, na.rm = T)

## [1] 23.62513

sd(dc$APPEARANCES, na.rm = T)

## [1] 87.37851

sum(is.na(dc$APPEARANCES)) # calcolo NA

## [1] 355
```

Var qualitative

```
table(dc$SEX)

##
##      Female Characters  Genderless Characters      Male Characters
##              1967              20              4783
## Transgender Characters
##              1
```

Summary

```
summary(dc)

##      page_id      name
##  Min.   : 1380   3g4 (New Earth)      : 1
##  1st Qu.: 44106  500-ZQ (New Earth)      : 1
##  Median :141267  Aa (New Earth)           : 1
##  Mean   :147441  Aarden (New Earth)       : 1
##  3rd Qu.:213203  Aaron Babcock (New Earth): 1
##  Max.   :404010  Aaron Cash (New Earth)   : 1
##              (Other)           :6890
##              urlslug           ID
##  \\wiki\\3g4_(New_Earth)       : 1  Identity Unknown: 9
##  \\wiki\\500-ZQ_(New_Earth)     : 1  Public Identity :2466
##  \\wiki\\A%27Hwiirdh-Paan%27A_(New_Earth): 1  Secret Identity :2408
##  \\wiki\\A%27monn_A%27mokk_(New_Earth)  : 1  NA's           :2013
```

```

##  \\/wiki\\/A%27morr_(New_Earth)      :    1
##  \\/wiki\\/Aa_(New_Earth)              :    1
##  (Other)                              :6890
##
##      ALIGN      EYE      HAIR
##  Bad Characters      :2895  Blue Eyes :1102  Black Hair:1574
##  Good Characters     :2832  Brown Eyes: 879  Brown Hair:1148
##  Neutral Characters: 565  Black Eyes: 412  Blond Hair: 744
##  Reformed Criminals:   3  Green Eyes: 291  Red Hair  : 461
##  NA's                : 601  Red Eyes  : 208  White Hair: 346
##                      (Other) : 376  (Other)   : 349
##                      NA's     :3628  NA's      :2274
##
##      SEX      GSM
##  Female Characters    :1967  Bisexual Characters : 10
##  Genderless Characters : 20  Homosexual Characters: 54
##  Male Characters      :4783  NA's              :6832
##  Transgender Characters: 1
##  NA's                 : 125
##
##
##      ALIVE      APPEARANCES      FIRST.APPEARANCE
##  Deceased Characters:1693  Min.   : 1.00  2010, December: 78
##  Living Characters  :5200  1st Qu.: 2.00  2006, June    : 48
##  NA's               :   3  Median : 6.00  1989, January : 45
##                      Mean   : 23.63  2009, October : 44
##                      3rd Qu.: 15.00  1988, March   : 40
##                      Max.   :3093.00 (Other)       :6572
##                      NA's   :355    NA's         : 69
##
##      YEAR
##  Min.   :1935
##  1st Qu.:1983
##  Median :1992
##  Mean   :1990
##  3rd Qu.:2003
##  Max.   :2013
##  NA's   :69

```