

Data Visualization

Honey, Vuthy, James, Sonya

March 24, 2015

Data Visualization for the IS607 Project No. 2 | CUNY MSDA

Load data

```
load("Awards.Data.RData")
```

```
#View(data.awards)
```

```
library(tidyr)
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(knitr)
```

```
#Library(grid)
```

```
library(gridExtra)
```

Table 1. Number of times a Best Picture winner was nominated or won another category.

```

# Create BP, a dataframe showing only Best Picture Winners
BP <- filter(data.awards, category_name == "BEST PICTURE", won == 1)

# Create BP.AlsoNom, a dataframe showing what other categories the Best Picture Winners were NOMINATE
D for
BP.AlsoNom <- left_join(BP, data.awards, by = "movie_name")
BP.AlsoNom <- filter(BP.AlsoNom, category_name.y != "BEST PICTURE")
BP.AlsoNom <- select(BP.AlsoNom, category_name.y)
BP.AlsoNom <- count(BP.AlsoNom, category_name.y)
BP.AlsoNom <- arrange(BP.AlsoNom, desc(n))

# Create BP.AlsoWon, a dataframe showing what other categories the Best Picture winners WON
data.awards.won <- filter(data.awards, won == 1)

BP.AlsoWon <- left_join(BP, data.awards.won, by = "movie_name")
BP.AlsoWon <- filter(BP.AlsoWon, category_name.y != "BEST PICTURE")
BP.AlsoWon <- select(BP.AlsoWon, category_name.y)
BP.AlsoWon <- count(BP.AlsoWon, category_name.y)
BP.AlsoWon <- arrange(BP.AlsoWon, desc(n))

kable(cbind(BP.AlsoNom, BP.AlsoWon),
      col.names = c("Category Nominations", "Count", "Category Winners", "Count"))

```

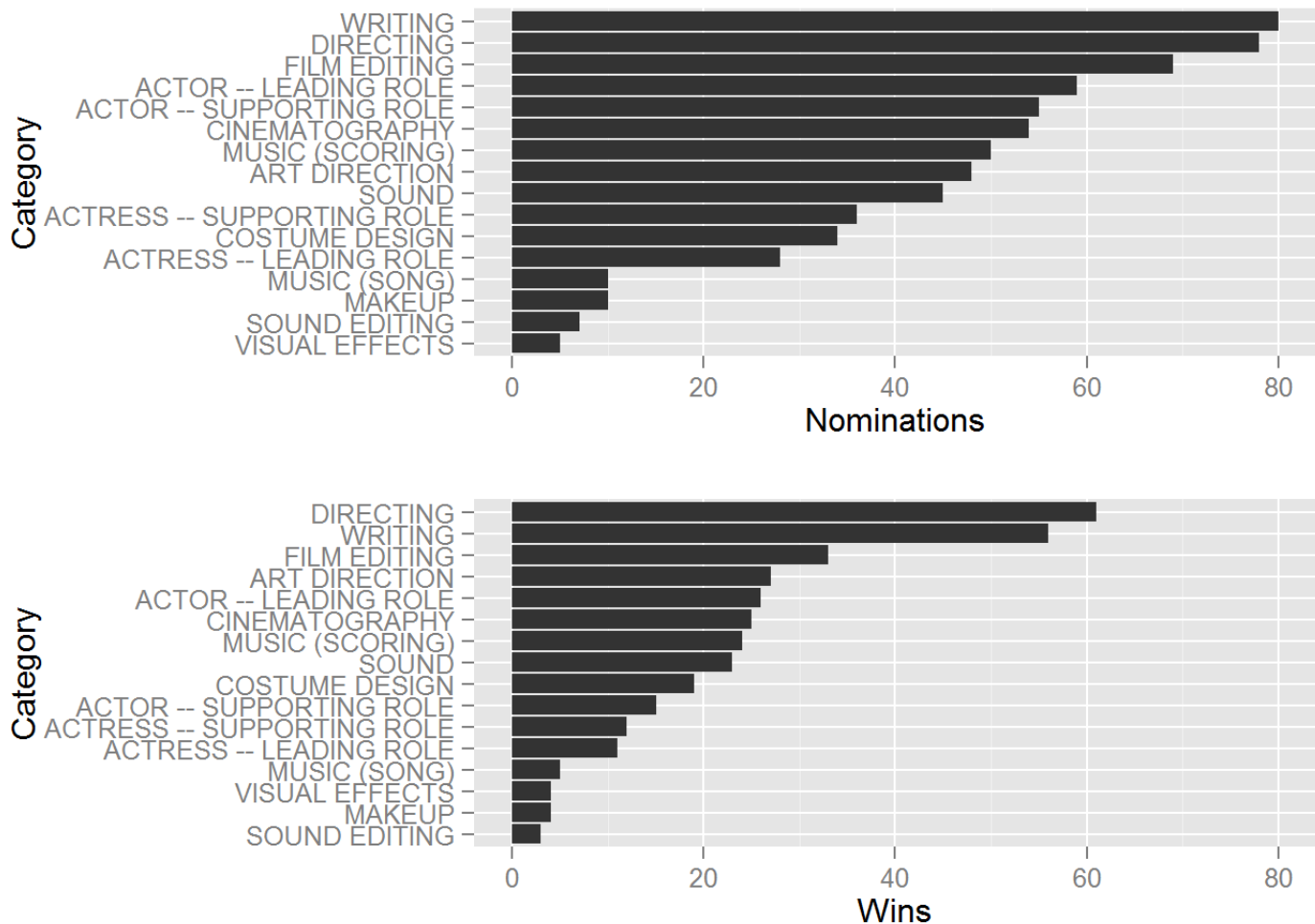
Category Nominations	Count	Category Winners	Count
WRITING	80	DIRECTING	61
DIRECTING	78	WRITING	56
FILM EDITING	69	FILM EDITING	33
ACTOR – LEADING ROLE	59	ART DIRECTION	27
ACTOR – SUPPORTING ROLE	55	ACTOR – LEADING ROLE	26
CINEMATOGRAPHY	54	CINEMATOGRAPHY	25
MUSIC (SCORING)	50	MUSIC (SCORING)	24
ART DIRECTION	48	SOUND	23
SOUND	45	COSTUME DESIGN	19
ACTRESS – SUPPORTING ROLE	36	ACTOR – SUPPORTING ROLE	15
COSTUME DESIGN	34	ACTRESS – SUPPORTING ROLE	12
ACTRESS – LEADING ROLE	28	ACTRESS – LEADING ROLE	11
MAKEUP	10	MUSIC (SONG)	5
MUSIC (SONG)	10	MAKEUP	4
SOUND EDITING	7	VISUAL EFFECTS	4

Figure 1. Barchart showing the number of times a Best Picture winner was nominated or won another category.

```
nom <- ggplot(data = BP.AlsoNom, aes(x = reorder(category_name.y, n), y=n)) + geom_bar(stat="identity") + coord_flip() + labs(x = "Category", y = "Nominations")

win <- ggplot(data = BP.AlsoWon, aes(x = reorder(category_name.y, n), y=n)) + geom_bar(stat="identity") + coord_flip() + labs(x = "Category", y = "Wins") + ylim(0, 80)

grid.arrange(nom, win, ncol=1)
```



Correlation Matrices

Use Sekhar's method to create award and category matrices

```
library(dplyr)
library(tidyr)

awards_df <- read.csv("Awards_File.csv",stringsAsFactors=F)
awards_modified <- awards_df[,c(1,3,4,6)]
head(awards_modified)
```

```
##   movie_id year category_id won
## 1      1 2010           1    0
## 2      2 2010           1    0
## 3      2 2010           4    0
## 4      2 2010           6    0
## 5      2 2010           7    0
## 6      2 2010           8    0
```

```
awards_modified <- awards_modified %>%
  group_by(movie_id,year,category_id) %>%
  summarise(won=max(won))
awards_re_modified <- spread((awards_modified),category_id,won)
head(awards_re_modified)
```

```
## Source: local data frame [6 x 25]
##
##   movie_id year  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
## 1      1 2010  0 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## 2      2 2010  0 NA NA  0 NA  0  0  0  0 NA NA NA NA NA NA  0 NA NA  0
## 3      3 2010  0 NA NA NA NA NA  0 NA  0 NA NA  1 NA  1 NA  0 NA NA  0
## 4      4 2010  1  0 NA  0 NA  0  0  0  1 NA NA  0 NA  0 NA  1 NA NA  0
## 5      5 2010  0 NA NA NA NA NA NA NA NA NA NA NA  0 NA  0  0  0 NA NA
## 6      6 2010 NA  1 NA  1 NA NA NA NA  0 NA NA  0 NA NA NA  0 NA NA NA
## Variables not shown: 20 (int), 21 (int), 22 (int), 23 (int)
```

```
award_categories <- unique(data.frame(category_id=awards_df$category_id,category_name=awards_df$cat
ory_name))
rownames(award_categories) <- NULL
names(awards_re_modified) <- c("movie_id", "year", "c1", "c2", "c3", "c4", "c5", "c6", "c7",
"c8", "c9", "c10", "c11", "c12", "c13", "c14", "c15","c16", "c17", "c18", "c19", "c20", "c21", "c2
2", "c23")
```

Modify awards matrix to set NA = 0

Figure 3. Correlation matrix, display ellipses

```
corrplot(W, method = "ellipse", tl.cex = 0.5)
```

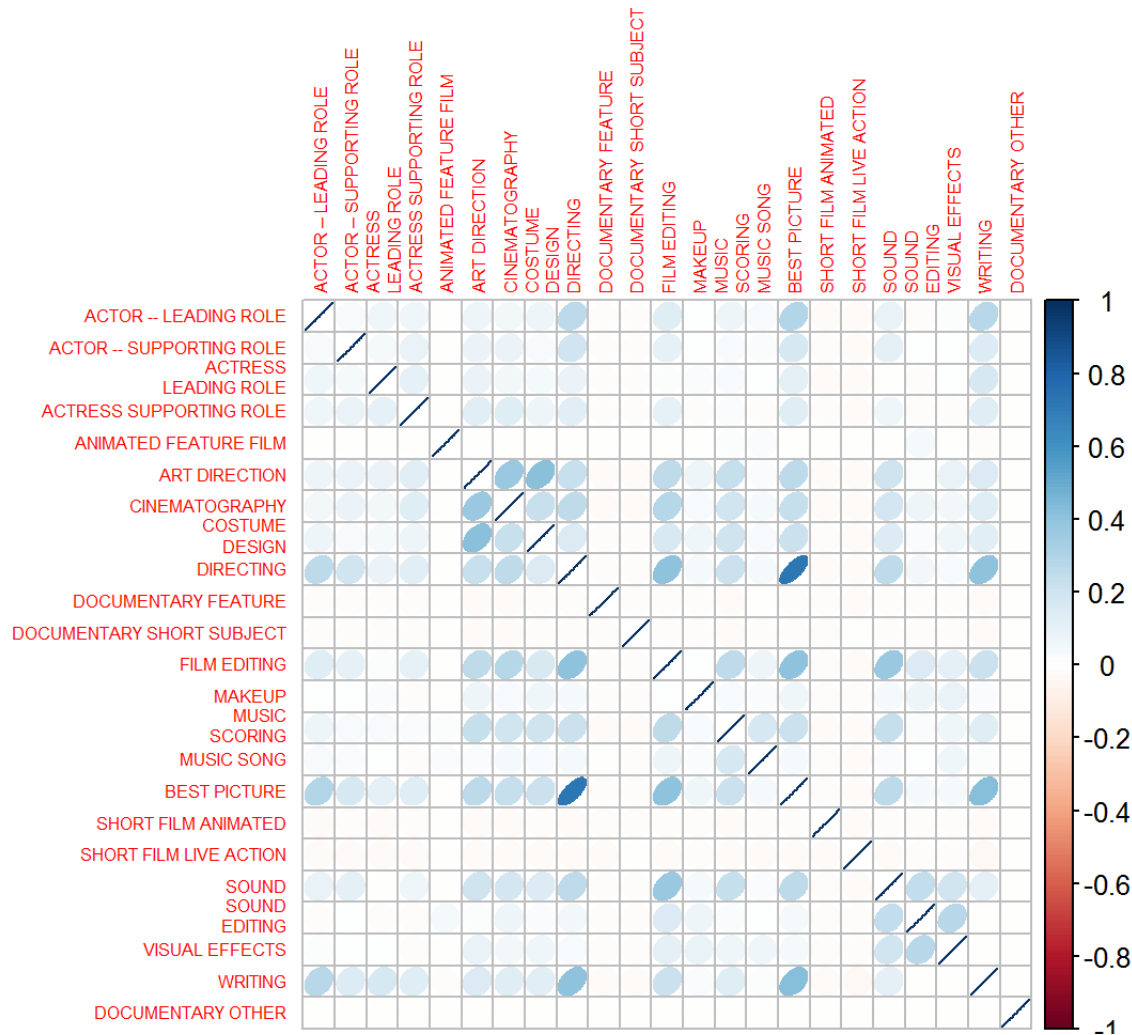


Figure 4. Correlation matrix, display squares

```
corrplot(W, method = "square", tl.cex = 0.5)
```

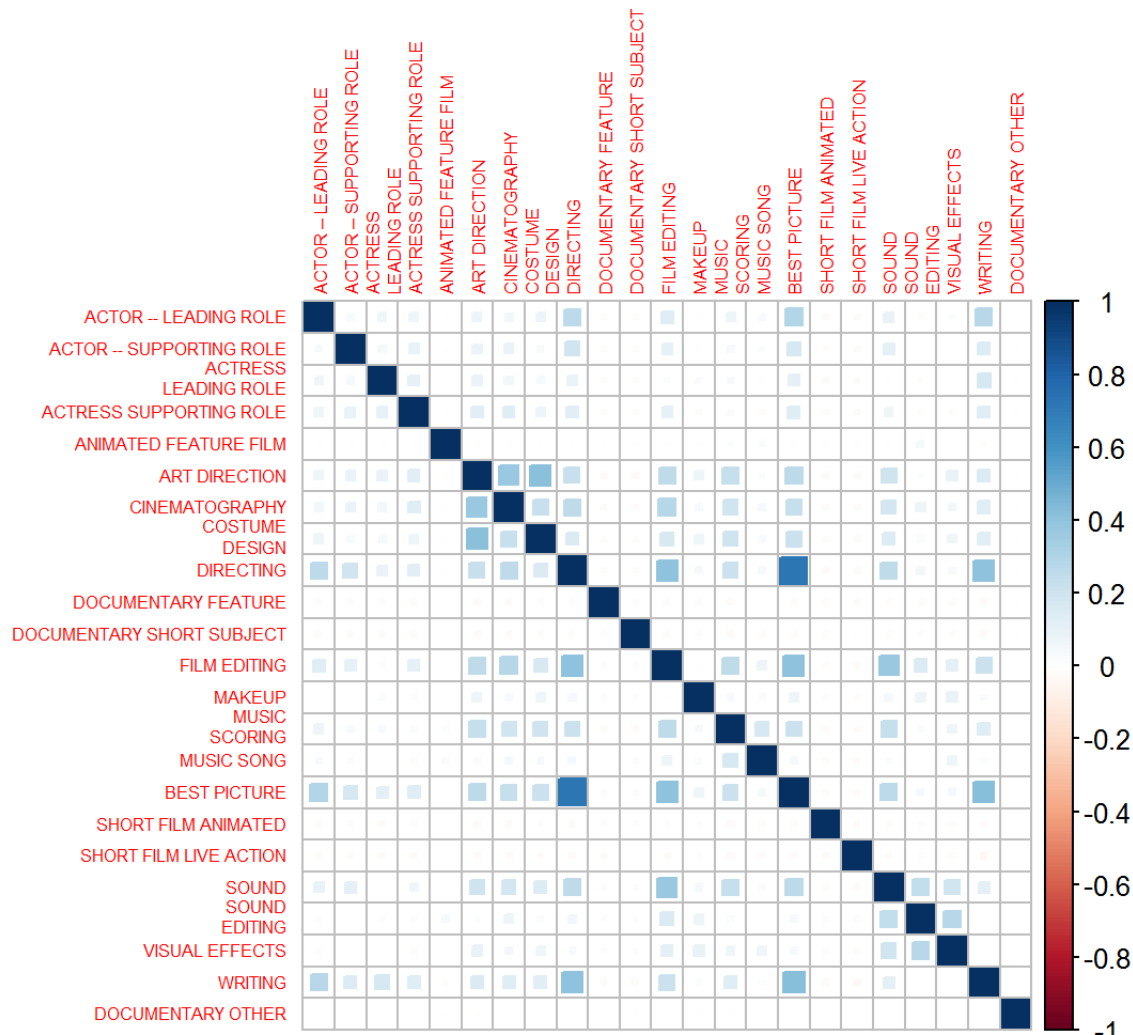


Figure 5. Correlation matrix, using ggplot2

```
library(ggplot2)
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.1.3
```

```
cm <- qplot(x=Var1, y=Var2, data=melt(cor(awards_mod[3:25])), fill=value, geom="tile")
cm + theme(axis.text=element_text(size=8))
```

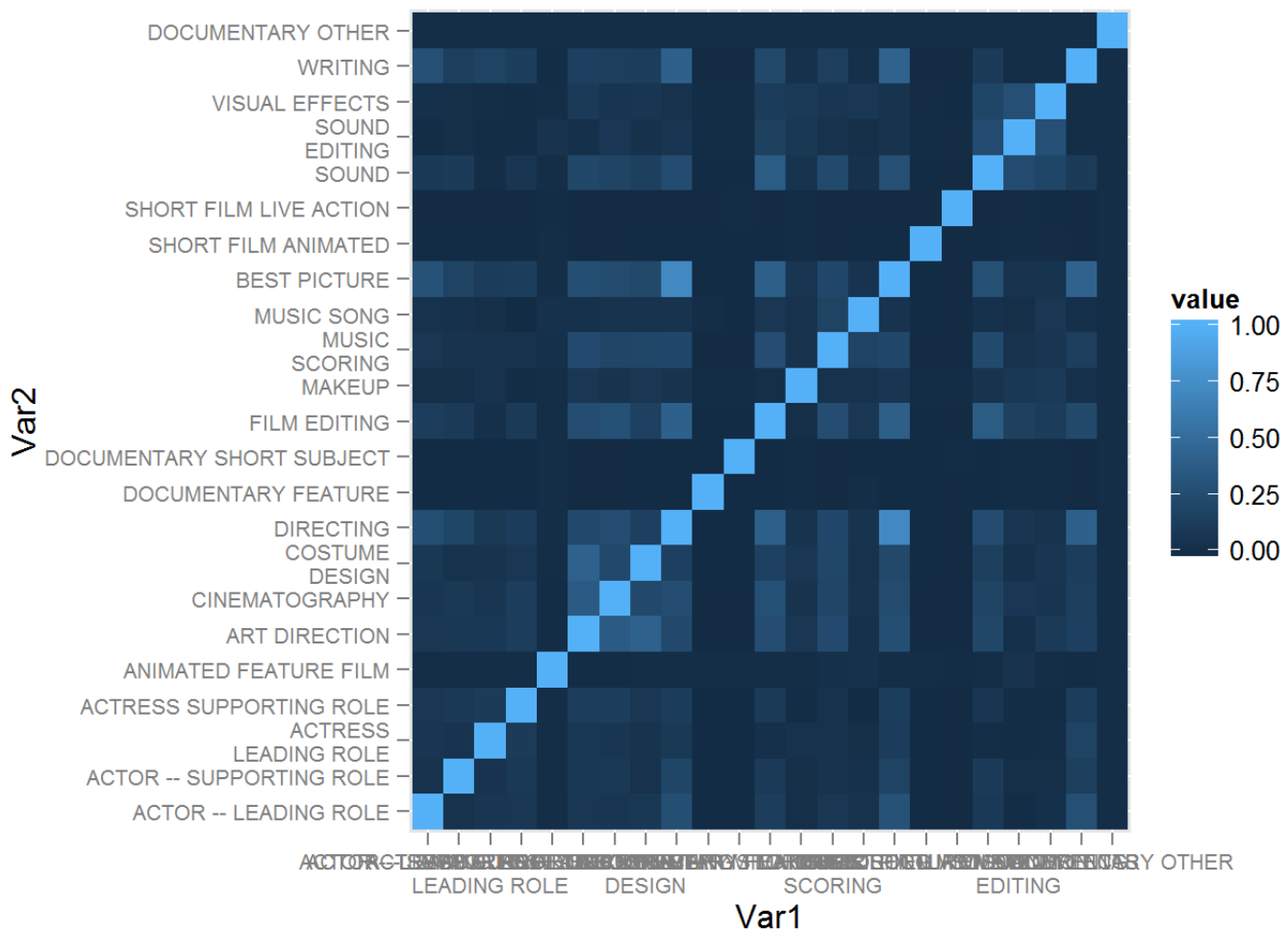


Table 2. Correlation Matrix, numerical values

```
x <- awards_mod[3:25]
y <- awards_mod[3:25]
cor(x, y)
```

##	ACTOR -- LEADING ROLE	ACTOR -- SUPPORTING ROLE
## ACTOR -- LEADING ROLE	1.00000000	0.031442183
## ACTOR -- SUPPORTING ROLE	0.031442183	1.00000000
## ACTRESS\nLEADING ROLE	0.063755763	0.044702504
## ACTRESS SUPPORTING ROLE	0.069755123	0.090204750
## ANIMATED FEATURE FILM	-0.007796584	-0.007314613
## ART DIRECTION	0.072506810	0.080054937
## CINEMATOGRAPHY	0.058936153	0.087843420
## COSTUME\nDESIGN	0.078686034	0.033580145
## DIRECTING	0.267328311	0.197464921
## DOCUMENTARY FEATURE	-0.018964075	-0.017791749
## DOCUMENTARY SHORT SUBJECT	-0.018825820	-0.017662041
## FILM EDITING	0.131378762	0.101967154
## MAKEUP	0.007197662	0.009224239
## MUSIC\nSCORING	0.074022300	0.039563592

## MUSIC SONG	0.030521724	0.021538257
## BEST PICTURE	0.295265477	0.174406734
## SHORT FILM ANIMATED	-0.020168626	-0.018921837
## SHORT FILM LIVE ACTION	-0.023553485	-0.022097450
## SOUND	0.091877626	0.113451180
## SOUND\nEDITING	-0.013917970	0.005933184
## VISUAL EFFECTS	0.016160545	0.002224411
## WRITING	0.272537746	0.141419198
## DOCUMENTARY OTHER	-0.004497010	-0.004219012
##	ACTRESS\nLEADING ROLE	ACTRESS SUPPORTING ROLE
## ACTOR -- LEADING ROLE	0.0637557625	0.069755123
## ACTOR -- SUPPORTING ROLE	0.0447025037	0.090204750
## ACTRESS\nLEADING ROLE	1.0000000000	0.108928458
## ACTRESS SUPPORTING ROLE	0.1089284579	1.0000000000
## ANIMATED FEATURE FILM	-0.0077496318	-0.007314613
## ART DIRECTION	0.0839561720	0.125503007
## CINEMATOGRAPHY	0.0595665577	0.132112232
## COSTUME\nDESIGN	0.0420486161	0.073047923
## DIRECTING	0.0879173749	0.120839042
## DOCUMENTARY FEATURE	-0.0188498701	-0.017791749
## DOCUMENTARY SHORT SUBJECT	-0.0187124480	-0.017662041
## FILM EDITING	0.0182542530	0.101967154
## MAKEUP	0.0272607454	-0.011783712
## MUSIC\nSCORING	0.0347082404	0.028980108
## MUSIC SONG	0.0055738644	-0.018676191
## BEST PICTURE	0.1138887590	0.135644487
## SHORT FILM ANIMATED	-0.0200471669	-0.018921837
## SHORT FILM LIVE ACTION	-0.0234116423	-0.022097450
## SOUND	-0.0075252400	0.060501973
## SOUND\nEDITING	-0.0138341537	-0.013057585
## VISUAL EFFECTS	0.0004362834	-0.014693358
## WRITING	0.1728635223	0.132467112
## DOCUMENTARY OTHER	-0.0044699280	-0.004219012
##	ANIMATED FEATURE FILM	ART DIRECTION
## ACTOR -- LEADING ROLE	-0.007796584	0.072506810
## ACTOR -- SUPPORTING ROLE	-0.007314613	0.080054937
## ACTRESS\nLEADING ROLE	-0.007749632	0.083956172
## ACTRESS SUPPORTING ROLE	-0.007314613	0.125503007
## ANIMATED FEATURE FILM	1.0000000000	-0.008813338
## ART DIRECTION	-0.008813338	1.0000000000
## CINEMATOGRAPHY	-0.009061580	0.370686282
## COSTUME\nDESIGN	-0.007559144	0.417374104
## DIRECTING	-0.007796584	0.222109306
## DOCUMENTARY FEATURE	-0.007062251	-0.021437183
## DOCUMENTARY SHORT SUBJECT	-0.007010764	-0.021280899
## FILM EDITING	-0.007413321	0.257903256
## MAKEUP	-0.004677423	0.073692036

## MUSIC\	SCORING	0.016754610	0.236824382
## MUSIC	SONG	0.025842324	0.022362133
## BEST	PICTURE	-0.007702417	0.268522271
## SHORT	FILM ANIMATED	-0.007510827	-0.022798820
## SHORT	FILM LIVE ACTION	-0.008771354	-0.026625099
## SOUND		-0.007510827	0.209798842
## SOUND\	EDITING	0.041930610	0.016047368
## VISUAL	EFFECTS	-0.005832377	0.095541216
## WRITING		-0.011380431	0.152717654
## DOCUMENTARY	OTHER	-0.001674693	-0.005083466
##	CINEMATOGRAPHY	COSTUME\	DESIGN DIRECTING
## ACTOR --	LEADING ROLE	0.05893615	0.078686034 0.267328311
## ACTOR --	SUPPORTING ROLE	0.08784342	0.033580145 0.197464921
## ACTRESS\	LEADING ROLE	0.05956656	0.042048616 0.087917375
## ACTRESS	SUPPORTING ROLE	0.13211223	0.073047923 0.120839042
## ANIMATED	FEATURE FILM	-0.00906158	-0.007559144 -0.007796584
## ART	DIRECTION	0.37068628	0.417374104 0.222109306
## CINEMATOGRAPHY		1.00000000	0.223022531 0.256699968
## COSTUME\	DESIGN	0.22302253	1.000000000 0.152924338
## DIRECTING		0.25669997	0.152924338 1.000000000
## DOCUMENTARY	FEATURE	-0.02204100	-0.018386537 -0.018964075
## DOCUMENTARY	SHORT SUBJECT	-0.02188031	-0.018252492 -0.018825820
## FILM	EDITING	0.28276926	0.162519375 0.408735615
## MAKEUP		0.03676774	0.069235822 0.046713306
## MUSIC\	SCORING	0.19461560	0.200844175 0.213373931
## MUSIC	SONG	0.04241460	0.032648003 0.043128854
## BEST	PICTURE	0.23922825	0.217791704 0.720580863
## SHORT	FILM ANIMATED	-0.02344099	-0.019554404 -0.020168626
## SHORT	FILM LIVE ACTION	-0.02737504	-0.022836180 -0.023553485
## SOUND		0.18154401	0.147168546 0.253722211
## SOUND\	EDITING	0.07669114	0.023303969 0.057524736
## VISUAL	EFFECTS	0.05073911	0.066768589 0.032071581
## WRITING		0.13958953	0.126487780 0.407247664
## DOCUMENTARY	OTHER	-0.00522665	-0.004360056 -0.004497010
##	DOCUMENTARY	FEATURE	DOCUMENTARY SHORT SUBJECT
## ACTOR --	LEADING ROLE	-0.018964075	-0.018825820
## ACTOR --	SUPPORTING ROLE	-0.017791749	-0.017662041
## ACTRESS\	LEADING ROLE	-0.018849870	-0.018712448
## ACTRESS	SUPPORTING ROLE	-0.017791749	-0.017662041
## ANIMATED	FEATURE FILM	-0.007062251	-0.007010764
## ART	DIRECTION	-0.021437183	-0.021280899
## CINEMATOGRAPHY		-0.022040997	-0.021880311
## COSTUME\	DESIGN	-0.018386537	-0.018252492
## DIRECTING		-0.018964075	-0.018825820
## DOCUMENTARY	FEATURE	1.000000000	-0.017052681
## DOCUMENTARY	SHORT SUBJECT	-0.017052681	1.000000000
## FILM	EDITING	-0.018031842	-0.017900384

## MAKEUP	-0.011377161	-0.011294218
## MUSIC\nSCORING	-0.023111450	-0.022942959
## MUSIC SONG	-0.004165055	-0.017900384
## BEST PICTURE	-0.018735026	-0.018598441
## SHORT FILM ANIMATED	-0.018269012	-0.018135825
## SHORT FILM LIVE ACTION	-0.021335064	-0.009288812
## SOUND	-0.018269012	-0.018135825
## SOUND\nEDITING	-0.012607084	-0.012515174
## VISUAL EFFECTS	-0.014186421	-0.014082997
## WRITING	-0.027681270	-0.027479464
## DOCUMENTARY OTHER	-0.004073452	-0.004043755
##	FILM EDITING	MAKEUP MUSIC\nSCORING
## ACTOR -- LEADING ROLE	0.131378762	0.007197662 0.07402230
## ACTOR -- SUPPORTING ROLE	0.101967154	0.009224239 0.03956359
## ACTRESS\nLEADING ROLE	0.018254253	0.027260745 0.03470824
## ACTRESS SUPPORTING ROLE	0.101967154	-0.011783712 0.02898011
## ANIMATED FEATURE FILM	-0.007413321	-0.004677423 0.01675461
## ART DIRECTION	0.257903256	0.073692036 0.23682438
## CINEMATOGRAPHY	0.282769264	0.036767744 0.19461560
## COSTUME\nDESIGN	0.162519375	0.069235822 0.20084417
## DIRECTING	0.408735615	0.046713306 0.21337393
## DOCUMENTARY FEATURE	-0.018031842	-0.011377161 -0.02311145
## DOCUMENTARY SHORT SUBJECT	-0.017900384	-0.011294218 -0.02294296
## FILM EDITING	1.000000000	0.008795693 0.25782754
## MAKEUP	0.008795693	1.000000000 0.03381370
## MUSIC\nSCORING	0.257827544	0.033813704 1.00000000
## MUSIC SONG	0.073701618	0.029534115 0.17424595
## BEST PICTURE	0.401247896	0.067549655 0.21648362
## SHORT FILM ANIMATED	-0.019177180	-0.012099810 -0.02457943
## SHORT FILM LIVE ACTION	-0.022395647	-0.014130497 -0.02870455
## SOUND	0.372846880	0.049337977 0.23334904
## SOUND\nEDITING	0.155490284	0.079791297 0.02744237
## VISUAL EFFECTS	0.118714091	0.095297214 0.06002750
## WRITING	0.218385155	0.023215303 0.13718831
## DOCUMENTARY OTHER	-0.004275947	-0.002697901 -0.00548049
##	MUSIC SONG	BEST PICTURE SHORT FILM ANIMATED
## ACTOR -- LEADING ROLE	0.030521724	0.295265477 -0.020168626
## ACTOR -- SUPPORTING ROLE	0.021538257	0.174406734 -0.018921837
## ACTRESS\nLEADING ROLE	0.005573864	0.113888759 -0.020047167
## ACTRESS SUPPORTING ROLE	-0.018676191	0.135644487 -0.018921837
## ANIMATED FEATURE FILM	0.025842324	-0.007702417 -0.007510827
## ART DIRECTION	0.022362133	0.268522271 -0.022798820
## CINEMATOGRAPHY	0.042414599	0.239228246 -0.023440987
## COSTUME\nDESIGN	0.032648003	0.217791704 -0.019554404
## DIRECTING	0.043128854	0.720580863 -0.020168626
## DOCUMENTARY FEATURE	-0.004165055	-0.018735026 -0.018269012
## DOCUMENTARY SHORT SUBJECT	-0.017900384	-0.018598441 -0.018135825

## FILM EDITING	0.073701618	0.401247896	-0.019177180
## MAKEUP	0.029534115	0.067549655	-0.012099810
## MUSIC\ nSCORING	0.174245949	0.216483617	-0.024579432
## MUSIC SONG	1.000000000	0.044108527	-0.019177180
## BEST PICTURE	0.044108527	1.000000000	-0.019925028
## SHORT FILM ANIMATED	-0.019177180	-0.019925028	1.000000000
## SHORT FILM LIVE ACTION	-0.022395647	-0.023269005	-0.022690214
## SOUND	0.020025226	0.269773396	-0.019429415
## SOUND\ nEDITING	0.005513327	0.040460535	-0.013407855
## VISUAL EFFECTS	0.068611942	0.048918152	-0.015087507
## WRITING	0.015128829	0.429787279	-0.029439516
## DOCUMENTARY OTHER	-0.004275947	-0.004442695	-0.004332187
##	SHORT FILM LIVE ACTION	SOUND	
## ACTOR -- LEADING ROLE	-0.023553485	0.091877626	
## ACTOR -- SUPPORTING ROLE	-0.022097450	0.113451180	
## ACTRESS\ nLEADING ROLE	-0.023411642	-0.007525240	
## ACTRESS SUPPORTING ROLE	-0.022097450	0.060501973	
## ANIMATED FEATURE FILM	-0.008771354	-0.007510827	
## ART DIRECTION	-0.026625099	0.209798842	
## CINEMATOGRAPHY	-0.027375040	0.181544012	
## COSTUME\ nDESIGN	-0.022836180	0.147168546	
## DIRECTING	-0.023553485	0.253722211	
## DOCUMENTARY FEATURE	-0.021335064	-0.018269012	
## DOCUMENTARY SHORT SUBJECT	-0.009288812	-0.018135825	
## FILM EDITING	-0.022395647	0.372846880	
## MAKEUP	-0.014130497	0.049337977	
## MUSIC\ nSCORING	-0.028704548	0.233349036	
## MUSIC SONG	-0.022395647	0.020025226	
## BEST PICTURE	-0.023269005	0.269773396	
## SHORT FILM ANIMATED	-0.022690214	-0.019429415	
## SHORT FILM LIVE ACTION	1.000000000	-0.022690214	
## SOUND	-0.022690214	1.000000000	
## SOUND\ nEDITING	-0.015658068	0.245771967	
## VISUAL EFFECTS	-0.017619613	0.199308678	
## WRITING	-0.034380290	0.110189225	
## DOCUMENTARY OTHER	-0.005059250	-0.004332187	
##	SOUND\ nEDITING	VISUAL EFFECTS	WRITING
## ACTOR -- LEADING ROLE	-0.013917970	0.0161605455	0.2725377457
## ACTOR -- SUPPORTING ROLE	0.005933184	0.0022244112	0.1414191978
## ACTRESS\ nLEADING ROLE	-0.013834154	0.0004362834	0.1728635223
## ACTRESS SUPPORTING ROLE	-0.013057585	-0.0146933581	0.1324671121
## ANIMATED FEATURE FILM	0.041930610	-0.0058323766	-0.0113804313
## ART DIRECTION	0.016047368	0.0955412164	0.1527176545
## CINEMATOGRAPHY	0.076691137	0.0507391100	0.1395895318
## COSTUME\ nDESIGN	0.023303969	0.0667685886	0.1264877804
## DIRECTING	0.057524736	0.0320715813	0.4072476641
## DOCUMENTARY FEATURE	-0.012607084	-0.0141864212	-0.0276812702

## DOCUMENTARY SHORT SUBJECT	-0.012515174	-0.0140829972	-0.0274794641
## FILM EDITING	0.155490284	0.1187140907	0.2183851548
## MAKEUP	0.079791297	0.0952972142	0.0232153029
## MUSIC\ nSCORING	0.027442375	0.0600275014	0.1371883115
## MUSIC SONG	0.005513327	0.0686119417	0.0151288288
## BEST PICTURE	0.040460535	0.0489181524	0.4297872785
## SHORT FILM ANIMATED	-0.013407855	-0.0150875074	-0.0294395156
## SHORT FILM LIVE ACTION	-0.015658068	-0.0176196131	-0.0343802898
## SOUND	0.245771967	0.1993086779	0.1101892247
## SOUND\ nEDITING	1.000000000	0.2735097458	-0.0077958126
## VISUAL EFFECTS	0.273509746	1.0000000000	-0.0005542952
## WRITING	-0.007795813	-0.0005542952	1.0000000000
## DOCUMENTARY OTHER	-0.002989557	-0.0033640699	-0.0065641452
##	DOCUMENTARY OTHER		
## ACTOR -- LEADING ROLE	-0.004497010		
## ACTOR -- SUPPORTING ROLE	-0.004219012		
## ACTRESS\ nLEADING ROLE	-0.004469928		
## ACTRESS SUPPORTING ROLE	-0.004219012		
## ANIMATED FEATURE FILM	-0.001674693		
## ART DIRECTION	-0.005083466		
## CINEMATOGRAPHY	-0.005226650		
## COSTUME\ nDESIGN	-0.004360056		
## DIRECTING	-0.004497010		
## DOCUMENTARY FEATURE	-0.004073452		
## DOCUMENTARY SHORT SUBJECT	-0.004043755		
## FILM EDITING	-0.004275947		
## MAKEUP	-0.002697901		
## MUSIC\ nSCORING	-0.005480490		
## MUSIC SONG	-0.004275947		
## BEST PICTURE	-0.004442695		
## SHORT FILM ANIMATED	-0.004332187		
## SHORT FILM LIVE ACTION	-0.005059250		
## SOUND	-0.004332187		
## SOUND\ nEDITING	-0.002989557		
## VISUAL EFFECTS	-0.003364070		
## WRITING	-0.006564145		
## DOCUMENTARY OTHER	1.000000000		

An analysis of the various correlation matrix visualizations and table of numerical values shows the primary determining factor is DIRECTING, followed by WRITING and then FILM EDITING.

Note: There was an issue with NA values; we could not get them to be ignored, so instead had to set them equal to 0, which may have resulted in some skewing of the results.