Introduction to Statistics Project

Project No. 4: Flags

Kuber Shahi

Professor Kumarjit Saha

2nd December 2019

Database: Flag Database		**Number of instances = 194**	**Number of attributes = 30**
SN.	**Attributes**	**Attributes Information**	**No. of missing values**
1	Country	Name of the countries	
2	Landmass	1 = N. America, 2 = S. America, 3 = Europe, 4 = Africa, 5 = Asia, 6 = Oceania	
3	Zone	Geographic quadrant, based on Greenwich and the Equator 1 = NE, 2 = SE, 3 = SW, 4 = NW	
4	Area	in thousands of square km	
5	Population	in round millions	1
6	Language	1 = English, 2 = Spanish, 3 = French, 4 = German, 5 = Slavic, 6 = Other Indo-European, 7 = Chinese, 8 = Arabic, 9 = Japanese/Turkish/Finnish/Magyar, 10 = Others	
7	Religion	0 = Catholic, 1 = Other Christian, 2 = Muslim, 3 = Buddhist, 4 = Hindu, 5 = Ethnic, 6 = Marxist, 7 = Others	1 (replaced with mean)
8	Bars	Number of vertical bars in the flag	
9	Stripes	Number of horizontal stripes in the flag	1 (replaced with mean)
10	Colours	Number of different colours in the flag	
11	Red	0 if red absent, 1 if red present in the flag	
12	Green	same for green	
13	Blue	same for blue	
14	Gold	same for gold (also yellow)	
15	White	same for white	
16	Black	same for black	
17	Orange	same for orange (also brown)	
18	Main Hue	predominant colour in the flag (tie-breaks decided by taking the topmost hue, if that fails then the most central hue, and if that fails the leftmost hue)	
19	Circles	Number of circles in the flag	
20	Crosses	Number of (upright) crosses	

21	Saltires	Number of diagonal crosses	
22	Quarters	Number of quartered sections	
23	Sunstars	Number of sun or star symbols	
24	Crescent	1 if a crescent moon symbol present, else 0	
25	Triangle	1 if any triangles present, 0 otherwise	
26	Icon	1 if an inanimate image present (e.g., a boat), otherwise 0	
27	Animate	1 if an animate image (e.g., an eagle, a tree, a human hand) present, 0 otherwise	
28	Text	1 if any letters or writing on the flag (e.g., a motto or slogan), 0 otherwise	
29	Topleft	colour in the top-left corner (moving right to decide tie-breaks)	
30	BottomRight	Colour in the bottom-left corner (moving left to decide tie-breaks)	

1) **Import the database into R**
   Database (flag.csv) imported to R with the variable name 'flagdata'.
   R command: > flagdata = read.csv("flag.csv", header = TRUE, na.strings = ("?"))

2) **Produce a contingency table between landmass and language. Is there a significant association ship between the two variables?**

   > tab1= table(flagdata$Landmass, flagdata$Language)

   > tab1

Language

Landmass	1	2	3	4	5	6	7	8	9	10
1	19	9	1	0	0	2	0	0	0	0
2	3	11	1	0	0	2	0	0	0	0
3	3	1	2	6	3	17	0	0	2	1
4	6	0	12	0	0	4	0	7	0	23
5	0	0	0	0	1	4	4	12	2	16
6	12	0	1	0	0	1	0	0	0	6

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test(tab1, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab1

X-squared = 286.09, df = NA, p-value = 0.0004998

At 5% significance level, the critical p-value is 0.05. Since 0.0004998 < 0.05, we reject null hypothesis. There is an association ship between Landmass and Language at 5% significance level. Also, the probability of X-squared value occurring greater than calculated 286.09 is 0.0004998 which is very small. This implies that given the two variables are independent, chances of the above observed values occurring is very less, and hence, there is a significant association ship between Landmass and Language.

3) Produce a contingency table between landmass and religion. Is there a significant association ship between the two variables?

> tab2= table(flagdata$Landmass, flagdata$Religion)

> tab2

Religion

Landmass	0	1	2	3	4	5	6	7
1	9	21	0	0	0	0	1	0
2	13	3	0	0	1	0	0	0
3	14	13	0	0	0	0	8	0
4	2	7	15	0	1	26	0	1
5	0	0	20	8	2	0	6	3
6	2	16	1	0	0	1	0	0

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test(tab1, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab2

X-squared = 272.45, df = NA, p-value = 0.0004998

At 5% significance level, the critical p-value is 0.05. Since 0.0004998 < 0.05, we reject null hypothesis. There is an association ship between Landmass and Religion at 5% significance level. Also, the probability of X-squared value occurring greater than calculated 276.09 is 0.0004998 which is very small. This implies that given the two variables are independent, chances of the above observed values occurring is very less, and hence, there is a significant association ship between Landmass and Religion.

4) Judge the association ship between religion and the variables listed in 8 − 27. Which of these variables are significantly associated and which one of the variables is most significantly associated with religion?

a. Religion and Bars

> tab3r8= table(flagdata$Religion, flagdata$Bars)

> tab3r8

Religion	Bars 0	1	2	3	5
0	27	1	3	9	0
1	55	1	1	2	1
2	29	2	1	4	0
3	7	0	1	0	0
4	4	0	0	0	0
5	20	2	1	4	0
6	13	0	0	2	0
7	4	0	0	0	0

As we can see though Bars is a discrete variable, there are only 5 levels of it, and we can treat Bars as a categorical variable with 5 levels.

Now, we can perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r8, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r8

X-squared = 23.6, df = NA, p-value = 0.5772

At 5% significance level, the critical p-value is 0.05. Since 0.5772 > 0.05, we accept null hypothesis. There is no association ship between Religion and Bars at 5% significance level.

b. Religion and Stripes

> tab3r9= table(flagdata$Religion, flagdata$Stripes)

> tab3r9

Stripes

Religion	0	1	2	3	4	5	6	7	9	11	13	14
0	17	0	5	15	0	2	0	0	1	0	0	0
1	44	4	2	6	0	2	0	0	1	0	1	0
2	20	0	4	11	0	0	0	0	0	0	0	1
3	6	0	1	0	0	1	0	0	0	0	0	0
4	2	0	0	1	1	0	0	0	0	0	0	0
5	12	0	3	5	0	4	1	1	0	1	0	0
6	7	0	1	5	0	2	0	0	0	0	0	0
7	2	0	1	0	0	1	0	0	0	0	0	0

As we can see though Stripes is a discrete variable, there are only 12 levels of it, and we can treat Stripes as a categorical variable with 12 levels.

Now, we can perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r9, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r9

X-squared = 118.06, df = NA, p-value = 0.07546

At 5% significance level, the critical p-value is 0.05. Since 0.07546> 0.05, we accept null hypothesis. There is no association ship between Religion and Stripes at 5% significance level.

c. Religion and Colours

> tab3r10= table(flagdata$Religion, flagdata$Colours)

> tab3r10

Colours

Religion	1	2	3	4	5	6	7	8
0	0	11	19	5	3	1	1	0
1	0	9	25	8	5	7	5	1
2	1	13	8	13	1	0	0	0
3	0	2	3	2	0	1	0	0
4	0	0	1	2	1	0	0	0
5	0	1	11	11	4	0	0	0
6	0	4	8	1	1	0	1	0
7	0	2	0	2	0	0	0	0

As we can see though Colours is a discrete variable, there are only 8 levels of it, and we can treat Colours as a categorical variable with 8 levels.

Now, we can perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r11, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r10

X-squared = 64.19, df = NA, p-value = 0.1389

At 5% significance level, the critical p-value is 0.05. Since 0.1389 > 0.05, we accept null hypothesis. There is no association ship between Religion and Colours at 5% significance level.

d. Religion and Red

> tab3r11= table(flagdata$Religion, flagdata$Red)

> tab3r11

Religion	Red 0	Red 1
0	10	30
1	14	46
2	9	27
3	1	7
4	2	2
5	4	23
6	0	15
7	1	3

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r11, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r11

X-squared = 7.9154, df = NA, p-value = 0.3493

At 5% significance level, the critical p-value is 0.05. Since 0.3493 > 0.05, we accept null hypothesis. There is no association ship between Religion and Red at 5% significance level.

e. Religion and Green

> tab3r12= table(flagdata$Religion, flagdata$Green)

> tab3r12

Religion	Green 0	Green 1
0	30	10
1	35	25
2	10	26
3	6	2
4	1	3
5	6	21
6	12	3
7	3	1

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r12, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r12

X-squared = 35.906, df = NA, p-value = 0.0004998

At 5% significance level, the critical p-value is 0.05. Since 0.0004998 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Green at 5% significance level.

f. Religion and Blue

> tab3r13= table(flagdata$Religion, flagdata$Blue)

> tab3r13

Blue

Religion	0	1
0	14	26
1	18	42
2	32	4
3	4	4
4	1	3
5	17	10
6	7	8
7	2	2

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r13, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r13

X-squared = 37.793, df = NA, p-value = 0.0004998

At 5% significance level, the critical p-value is 0.05. Since 0.0004998 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Blue at 5% significance level.

g. Religion and Gold

> tab3r14= table(flagdata$Religion, flagdata$Gold)

> tab3r14

Gold

Religion	0	1
0	22	18
1	25	35
2	28	8
3	5	3
4	2	2
5	12	15
6	6	9
7	3	1

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r14, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r14

X-squared = 14.926, df = NA, p-value = 0.02549

At 5% significance level, the critical p-value is 0.05. Since 0.02549 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Gold at 5% significance level.

h. Religion and White

> tab3r15= table(flagdata$Religion, flagdata$White)

> tab3r15

Religion	White 0	White 1
0	22	18
1	25	35
2	28	8
3	5	3
4	2	2
5	12	15
6	6	9
7	3	1

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r15, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r15

X-squared = 14.737, df = NA, p-value = 0.03248

At 5% significance level, the critical p-value is 0.05. Since 0.03248 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and White at 5% significance level.

i. Religion and Black

> tab3r16= table(flagdata$Religion, flagdata$Black)

> tab3r16

Religion	Black 0	Black 1
0	31	9
1	46	14
2	25	11
3	7	1
4	3	1
5	14	13
6	13	2
7	3	1

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r16, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r16

X-squared = 9.5087, df = NA, p-value = 0.2154

At 5% significance level, the critical p-value is 0.05. Since 0.2154 > 0.05, we accept null hypothesis. There is no association ship between Religion and Black at 5% significance level.

j. Religion and Orange

> tab3r17= table(flagdata$Religion, flagdata$Orange)

> tab3r17

Religion	Orange 0	Orange 1
0	36	4
1	49	11
2	33	3
3	5	3
4	2	2
5	25	2
6	14	1
7	4	0

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r17, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r17

X-squared = 13.114, df = NA, p-value = 0.07396

At 5% significance level, the critical p-value is 0.05. Since 0.07396 > 0.05, we accept null hypothesis. There is no association ship between Religion and Orange at 5% significance level.

k. Religion and Main Hue

> tab3r18= table(flagdata$Religion, flagdata$Orange)

> tab3r18

Main Hue

Religion	black	blue	brown	gold	green	orange	red	white
0	1	9	0	6	3	0	15	6
1	1	24	0	3	6	1	16	9
2	2	2	1	3	12	1	15	0
3	0	1	0	1	0	1	4	1
4	0	0	1	0	1	1	1	0
5	1	2	0	5	8	0	10	1
6	0	2	0	1	0	0	10	2
7	0	0	0	0	1	0	0	3

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r18, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r18

X-squared = 115.09, df = NA, p-value = 0.0009995

At 5% significance level, the critical p-value is 0.05. Since 0.0009995 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Main Hue at 5% significance level.

l. Religion and Circles

> tab3r19= table(flagdata$Religion, flagdata$Circles)

> tab3r19

Circles

Religion	0	1	2	4
0	36	4	0	0
1	51	9	0	0
2	33	3	0	0
3	5	2	0	1
4	3	1	0	0
5	23	4	0	0
6	12	2	1	0
7	2	2	0	0

As we can see though Circles is a discrete variable, there are only 4 levels of it, and we can treat Circles as a categorical variable with 4 levels.

Now, we can perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r19, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r19

X-squared = 42.751, df = NA, p-value = 0.08946

At 5% significance level, the critical p-value is 0.05. Since 0.08946 > 0.05, we accept null hypothesis. There is no significant association ship between Religion and Circles at 5% significance level.

m. Religion and Crosses

> tab3r20= table(flagdata$Religion, flagdata$Crosses)

> tab3r20

Crosses

Religion	0	1	2
0	38	2	0
1	36	22	2
2	36	0	0
3	7	1	0
4	4	0	0
5	27	0	0
6	15	0	0
7	4	0	0

As we can see though Crosses is a discrete variable, there are only 3 levels of it, and we can treat Crosses as a categorical variable with 3 levels.

Now, we will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r20, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r20

X-squared = 50.858, df = NA, p-value = 0.002999

At 5% significance level, the critical p-value is 0.05. Since 0.002999 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Crosses at 5% significance level.

n. Religion and Saltires

> tab3r21= table(flagdata$Religion, flagdata$Saltires)

> tab3r21

Saltires

Religion	0	1
0	40	0
1	44	16
2	36	0
3	7	1
4	4	0
5	26	1
6	15	0
7	4	0

As we can see though Saltires is a discrete variable, there are only 2 levels of it, and we can treat Saltires as a categorical variable with 2 levels.

Now, we can perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r21, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r21

X-squared = 32.772, df = NA, p-value = 0.0009995

At 5% significance level, the critical p-value is 0.05. Since 0.0009995 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Saltires at 5% significance level.

o. Religion and Quarters

> tab3r22= table(flagdata$Religion, flagdata$Quarters)

> tab3r22

Quarters

Religion	0	1	4
0	37	2	1
1	43	17	0
2	35	1	0
3	5	3	0
4	4	0	0
5	26	1	0
6	15	0	0
7	3	1	0

As we can see though Quarters is a discrete variable, there are only 3 levels of it, and we can treat Quarters as a categorical variable with 3 levels.

Now, we will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r22, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r22

X-squared = 31.637, df = NA, p-value = 0.05747

At 5% significance level, the critical p-value is 0.05. Since 0.05747 > 0.05, we accept null hypothesis. There is no significant association ship between Religion and Quarters at 5% significance level.

p. Religion and Sunstars

> tab3r23= table(flagdata$Religion, flagdata$Sunstars)

> tab3r23

Sunstars

Religion	0	1	2	3	4	5	6	7	9	10	14	15	22	50
0	27	7	2	0	1	1	0	1	0	0	0	0	1	0
1	40	6	1	0	3	3	2	1	1	1	0	1	0	1
2	19	14	1	1	1	0	0	0	0	0	0	0	0	0
3	5	1	0	0	0	1	0	0	0	0	1	0	0	0
4	3	0	1	0	0	0	0	0	0	0	0	0	0	0
5	14	11	0	1	0	1	0	0	0	0	0	0	0	0
6	5	8	1	0	0	1	0	0	0	0	0	0	0	0
7	1	3	0	0	0	0	0	0	0	0	0	0	0	0

As we can see though Sunstars is a discrete variable, there are only 14 values of it, and we can treat Sunstars as a categorical variable with 14 levels.

Now, we can perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r23, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r23

X-squared = 88.379, df = NA, p-value = 0.4558

At 5% significance level, the critical p-value is 0.05. Since 0.4558 > 0.05, we accept null hypothesis. There is no significant association ship between Religion and Sunstars at 5% significance level.

q. Religion and Crescent

> tab3r24= table(flagdata$Religion, flagdata$Crescent)

> tab3r24

Crescents

Religion	0	1
0	40	0
1	60	0
2	28	8
3	7	1
4	3	1
5	27	0
6	14	1
7	4	0

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r24, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r24

X-squared = 29.835, df = NA, p-value = 0.002999

At 5% significance level, the critical p-value is 0.05. Since 0.002999 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Crescents at 5% significance level.

r. Religion and Triangle

> tab3r25= table(flagdata$Religion, flagdata$Triangle)

> tab3r25

Triangle

Religion	0	1
0	37	3
1	50	10
2	31	5
3	8	0
4	3	1
5	22	5
6	12	3
7	4	0

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r25, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r25

X-squared = 5.0441, df = NA, p-value = 0.6662

At 5% significance level, the critical p-value is 0.05. Since 0.6662 > 0.05, we accept null hypothesis. There is no significant association ship between Religion and Triangle at 5% significance level.

s. Religion and Icon

> tab3r26= table(flagdata$Religion, flagdata$Icon)

> tab3r26

Religion	Icon 0	Icon 1
0	33	7
1	41	19
2	31	5
3	4	4
4	3	1
5	20	7
6	10	5
7	3	1

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since some of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r26, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r26

X-squared = 8.1645, df = NA, p-value = 0.3058

At 5% significance level, the critical p-value is 0.05. Since 0.3058 > 0.05, we accept null hypothesis. There is no significant association ship between Religion and Icon at 5% significance level.

t. Religion and Animate

> tab3r27= table(flagdata$Religion, flagdata$Animate)

> tab3r27

Religion	Animate 0	Animate 1
0	36	4
1	41	19
2	33	3
3	4	4
4	4	0
5	21	6
6	12	3
7	4	0

We will perform chi-square test on it to check whether there is an association between the two variables.

Chi-square Test

H_0 : There two variables are independent of each other.

H_1 : There two variables are not independent of each other.

Significance Level: 5 %

Since many of the table values are less than 5, we will use Monte Carlo simulation to reduce the error.

> chisq.test (tab3r27, simulate.p.value=TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tab3r27

X-squared = 17.182, df = NA, p-value = 0.01599

At 5% significance level, the critical p-value is 0.05. Since 0.01599 < 0.05, we reject null hypothesis. There is a significant association ship between Religion and Animate at 5% significance level.

Now, to check which of the variables from 8-27 is most significantly associated with Religion, we look at the variables which does have association ship with Religion i.e. in our case, the tests where null hypothesis were rejected. Chi-square test only tells us whether there is an association ship between two categorical variable or not. So, to measure the strength of the strength of the association ship, we are going to use Cramer's V which builds upon the chi-square statistic value.

Cramer's V is computed by taking the square root of the chi-squared statistic divided by the sample size and the minimum dimension minus 1:

$$V = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}}$$

where χ^2 is from the Pearson's chi-squared test, n is the total observations, k is the number of columns, and r the number of rows. The p-value for the significance of V is the same one that is calculated using the Pearson's chi-squared test.

Cramer's V varies from 0 (implying no association between the variables) to 1 (complete association).

Hence, the Cramer V for the tests where null hypothesis was rejected:

e) Religion and Green:

> cramerV(tab3r12)

Cramer V

0.4302

f) Religion and Blue:

> cramerV(tab3r13)

Cramer V

0.4414

g) Religion and Gold:

> cramerV(tab3r14)

Cramer V

0.2774

h) Religion and White:

> cramerV(tab3r15)

Cramer V

0.2756

i) Religion and Main Hue:

> cramerV(tab3r18)

Cramer V

0.2911

m) Religion and Crosses:

> cramerV(tab3r20)

Cramer V

0.362

n) Religion and Saltires:

> cramerV(tab3r21)

Cramer V

0.411

q) Religion and Crescents:

> cramerV(tab3r24)

Cramer V

0.3922

t) Religion and Animate:

```
> cramerV(tab3r27)
Cramer V
0.2976
```

As we can see Religion and Blue association ship has the highest Cramer's V value of 0.4414. Hence, the categorical variable Blue is most significantly associated with Religion among all the variables from 8-27.