

R - Projekt

Highest Mountains

Iva Bokšić

22.1.2021.

Učitavanje dataframe u varijablu data.

```
data <- read.csv("HighestMountains.csv",encoding = 'UTF-8')
head(data)
```

```
## Rank Mountain Height..m. Height..ft.
## 1 1 Mount Everest / Sagarmatha / Chomolungma 8848 29029
## 2 2 K2 / Qogir / Godwin Austen 8611 28251
## 3 3 Kangchenjunga 8586 28169
## 4 4 Lhotse 8516 27940
## 5 5 Makalu 8485 27838
## 6 6 Cho Oyu 8188 26864
## Prominence..m. Range Coordinates
## 1 8848 Mahalangur Himalaya 27°59'17"N 86°55'31"E<U+FEFF>
## 2 4017 Baltoro Karakoram 35°52'53"N 76°30'48"E<U+FEFF>
## 3 3922 Kangchenjunga Himalaya 27°42'12"N 88°08'51"E<U+FEFF>
## 4 610 Mahalangur Himalaya 27°57'42"N 86°55'59"E<U+FEFF>
## 5 2386 Mahalangur Himalaya 27°53'23"N 87°05'20"E<U+FEFF>
## 6 2340 Mahalangur Himalaya 28°05'39"N 86°39'39"E<U+FEFF>
## Parent.mountain First.ascent Ascents.bef..2004 Failed.attempts.bef..2004
## 1 1953 >>145 121
## 2 Mount Everest 1954 45 44
## 3 Mount Everest 1955 38 24
## 4 Mount Everest 1956 26 26
## 5 Mount Everest 1955 45 52
## 6 Mount Everest 1954 79 28
```

Podaci dataframe-a su podijeljeni u 118 redaka i 11 stupaca.

```
dim(data)
```

```
## [1] 118 11
```

Nazivi stupaca su :

```
names(data)
```

```
## [1] "Rank" "Mountain"
## [3] "Height..m." "Height..ft."
## [5] "Prominence..m." "Range"
## [7] "Coordinates" "Parent.mountain"
## [9] "First.ascent" "Ascents.bef..2004"
## [11] "Failed.attempts.bef..2004"
```

Vidimo da se nazivi stupaca razlikuju u odnosu na csv dokument.

Preimenovanje stupaca da odgovara csv-u.

```
names(data)[names(data) == "Failed.attempts.bef..2004"] <- "Failed attempts bef. 2004."
names(data)[names(data) == "Height..m."] <- "Height (m)"
names(data)[names(data) == "Height..ft."] <- "Height (ft)"
names(data)[names(data) == "Prominence..m."] <- "Prominence (m)"
names(data)[names(data) == "Parent.mountain"] <- "Parent mountain"
names(data)[names(data) == "First.ascent"] <- "First ascent"
names(data)[names(data) == "Ascents.bef..2004"] <- "Ascents bef. 2004."
```

Nova imena stupaca su:

```
names(data)
```

```
## [1] "Rank" "Mountain"
## [3] "Height (m)" "Height (ft)"
## [5] "Prominence (m)" "Range"
## [7] "Coordinates" "Parent mountain"
## [9] "First ascent" "Ascents bef. 2004."
## [11] "Failed attempts bef. 2004."
```

##Opis dataframa

Rank

- Pozicija planinskog vrha s obzirom na visinu.

Mountain

- Ime planinskog vrha.
- U slučaju više imena međusobno su odvojeni s “/”.

Height (m)

- Visina planinskog vrha izražena u metrima.

Height (ft)

- Visina planinskog vrha izražena u stopama.

Prominence (m)

- Prominencija planine izražena u metrima

Range

- Planinski lanac kojem planinski vrh pripada.

Coordinates

- Koordinate vrha.

Parent mountain

- Roditeljska planina

First ascent

- Godina prvog penjanja na vrha.

Ascents bef. 2004

- Broj uspješnih penjanja prije 2004. godine.

Failed attempts bef. 2004:

- Broj neuspješnih penjanja prije 2004. godine.

Struktura dataframe je:

```
str(data)
```

```
## 'data.frame':   118 obs. of  11 variables:
## $ Rank          : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Mountain      : chr  "Mount Everest / Sagarmatha / Chomolungma" "K2 / Qogir / Godwin A." ...
## $ Height (m)     : int  8848 8611 8586 8516 8485 8188 8167 8163 8126 8091 ...
## $ Height (ft)    : int  29029 28251 28169 27940 27838 26864 26795 26781 26660 26545 ...
## $ Prominence (m) : int  8848 4017 3922 610 2386 2340 3357 3092 4608 2984 ...
## $ Range          : chr  "Mahalangur Himalaya" "Baltoro Karakoram" "Kangchenjunga Himalaya" ...
## $ Coordinates    : chr  "27°59'17\"N 86°55'31\"E<U+FEFF> " "35°52'53\"N 76°30'48\"E<U+FEFF> " ...
## $ Parent mountain : chr  "" "Mount Everest" "Mount Everest" "Mount Everest" ...
## $ First ascent   : chr  "1953" "1954" "1955" "1956" ...
## $ Ascents bef. 2004. : chr  ">>145 " "45" "38" "26" ...
## $ Failed attempts bef. 2004.: int  121 44 24 26 52 28 39 45 67 47 ...
```

Stupac Ascents.bef..2004 nije Integer i moramo ga pretvoriti da bi mogli raditi analize vezane za broj uspješnih i neuspješnih penjanja.

```
data$`Ascents bef. 2004.` <- as.numeric(data$`Ascents bef. 2004.`)
```

```
## Warning: NAs introduced by coercion
```

Dataframe sadrži 4 NA vrijednosti.

```
sum(!complete.cases(data))
```

```
## [1] 4
```

Prilikom pretvrobe vidimo da su pojednini stupci NA vrijednost, jer nisu sadržavali podatak o uspješnim penjanjima ili je taj podatak bio neprecizan. Takve planinske vrhove nećemo koristiti pri analizi vezanoj za broj uspješnih i neuspješnih penjanja.

```
na <- data[is.na(data$`Ascents bef. 2004.`),]
na
```

```
##      Rank                      Mountain Height (m) Height (ft)
## 1      1 Mount Everest / Sagarmatha / Chomolungma      8848      29029
## 48    43                      Muztagh Ata              7546      24757
## 55    50                      Ismoil Somoni Peak        7495      24590
## 65    60      Jengish Chokusu / Tömür / Pk Pobeda        7439      24406
##      Prominence (m)                      Range              Coordinates
## 1              8848      Mahalangur Himalaya 27°59'17"N 86°55'31"E<U+FEFF>
## 48              2735      Muztagata (Eastern Pamirs) 38°16'33"N 75°06'58"E<U+FEFF>
## 55              3402 Pamir (Akademiya Nauk Range) 38°56'35"N 72°00'57"E<U+FEFF>
## 65              4148      Tian Shan 42°02'05"N 80°07'47"E<U+FEFF>
##      Parent mountain First ascent Ascents bef. 2004.
## 1              1953              NA
## 48      Kongur Tagh      1956              NA
## 55      Muztagh Ata      1933              NA
## 65 Ismail Samani Peak      1938              NA
##      Failed attempts bef. 2004.
## 1              121
## 48              NA
## 55              NA
## 65              NA
```

Izbacivanje NA vrijednosti i kreiranje novog dataframe.

```
dataAnalyze <- data
dataAnalyze <- dataAnalyze[!dataAnalyze$Rank==1 &
                           !dataAnalyze$Rank==43 &
                           !dataAnalyze$Rank==50 &
                           !dataAnalyze$Rank==60,]
head(dataAnalyze)
```

```
##      Rank                      Mountain Height (m) Height (ft) Prominence (m)
## 2      2 K2 / Qogir / Godwin Austen      8611      28251      4017
## 3      3      Kangchenjunga              8586      28169      3922
## 4      4      Lhotse                    8516      27940       610
## 5      5      Makalu                    8485      27838      2386
## 6      6      Cho Oyu                   8188      26864      2340
## 7      7      Dhaulagiri I              8167      26795      3357
##      Range              Coordinates Parent mountain
## 2      Baltoro Karakoram 35°52'53"N 76°30'48"E<U+FEFF> Mount Everest
## 3 Kangchenjunga Himalaya 27°42'12"N 88°08'51"E<U+FEFF> Mount Everest
## 4      Mahalangur Himalaya 27°57'42"N 86°55'59"E<U+FEFF> Mount Everest
```

```
## 5    Mahalangur Himalaya 27°53'23"N 87°05'20"E<U+FEFF>    Mount Everest
## 6    Mahalangur Himalaya 28°05'39"N 86°39'39"E<U+FEFF>    Mount Everest
## 7    Dhaulagiri Himalaya 28°41'48"N 83°29'35"E<U+FEFF>      K2
##    First ascent Ascents bef. 2004. Failed attempts bef. 2004.
## 2          1954          45          44
## 3          1955          38          24
## 4          1956          26          26
## 5          1955          45          52
## 6          1954          79          28
## 7          1960          51          39
```

```
str(dataAnalyze)
```

```
## 'data.frame':    114 obs. of  11 variables:
## $ Rank           : int  2 3 4 5 6 7 8 9 10 11 ...
## $ Mountain       : chr  "K2 / Qogir / Godwin Austen" "Kangchenjunga" "Lhotse" "Makalu" .
## $ Height (m)      : int  8611 8586 8516 8485 8188 8167 8163 8126 8091 8080 ...
## $ Height (ft)     : int  28251 28169 27940 27838 26864 26795 26781 26660 26545 26509 ...
## $ Prominence (m)  : int  4017 3922 610 2386 2340 3357 3092 4608 2984 2155 ...
## $ Range           : chr  "Baltoro Karakoram" "Kangchenjunga Himalaya" "Mahalangur Himalaya"
## $ Coordinates     : chr  "35°52'53\"N 76°30'48\"E<U+FEFF> " "27°42'12\"N 88°08'51\"E<U+FEFF>
## $ Parent mountain : chr  "Mount Everest" "Mount Everest" "Mount Everest" "Mount Everest"
## $ First ascent    : chr  "1954" "1955" "1956" "1955" ...
## $ Ascents bef. 2004. : num  45 38 26 45 79 51 49 52 36 31 ...
## $ Failed attempts bef. 2004.: int  44 24 26 52 28 39 45 67 47 16 ...
```

Novi dataframe **dataAnalyze** ne sadržava NA vrijednosti.

```
sum(!complete.cases(dataAnalyze))
```

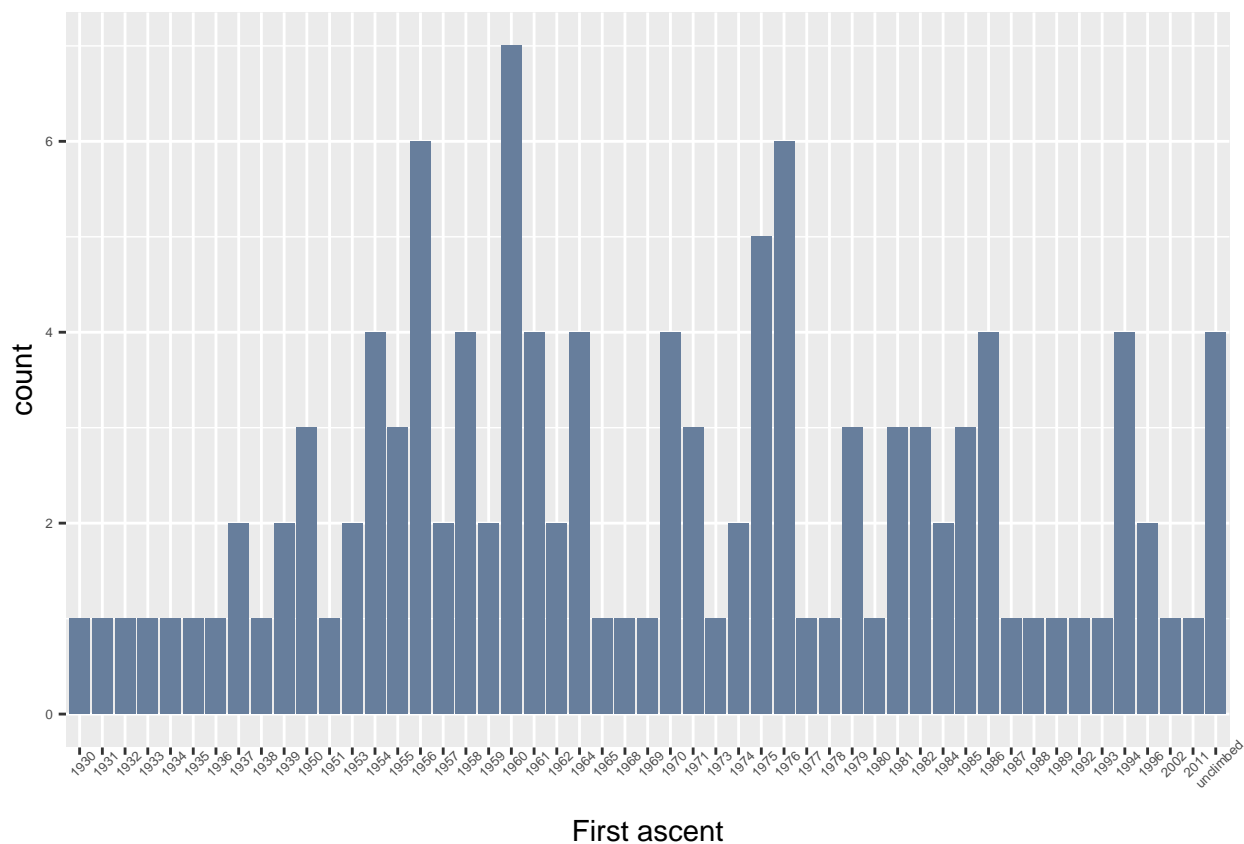
```
## [1] 0
```

Planinski lanci na kojima se nalaze najviše planine su :

```
unique(data$Range)
```

```
## [1] "Mahalangur Himalaya"      "Baltoro Karakoram"
## [3] "Kangchenjunga Himalaya"   "Dhaulagiri Himalaya"
## [5] "Manaslu Himalaya"        "Nanga Parbat Himalaya"
## [7] "Annapurna Himalaya"      "Jugal Himalaya"
## [9] "Hispar Karakoram"        "Masherbrum Karakoram"
## [11] "Garhwal Himalaya"        "Batura Karakoram"
## [13] "Rakaposhi-Haramosh Karakoram" "Assam Himalaya"
## [15] "Saltoro Karakoram"       "Hindu Kush"
## [17] "Langtang Himalaya"       "Nalakankar Himalaya"
## [19] "Saser Karakoram"         "Kongur Shan (Eastern Pamirs)"
## [21] "Kula Kangri Himalaya"    "Daxue Shan (Hengduan Shan)"
## [23] "Muztagata (Eastern Pamirs)" "Rimo Karakoram"
## [25] "Pamir (Akademiya Nauk Range)" "Siachen Karakoram"
## [27] "Tian Shan"              "Ganesh Himalaya"
## [29] "Labuche Himalaya"       "Jomolhari Himalaya"
## [31] "Yengisogat Karakoram"    "Panmah Karakoram"
## [33] "Baiku Himalaya"         "Lunana Himalaya"
## [35] "Nagarze Himalaya"
```

```
ggplot(data,aes(x=`First ascent`))+geom_bar(fill = "#677e9c")+theme(axis.text=element_text(size=5),axis
```



4 planinska vrha do 2004 su ostala neosvojena. Vidimo da su svi iznad 7000 metra i da pripadaju lancima Kula Kangri Himalaya i Lunana Himalaya.

```
unclimbed <- data[data$`First ascent`=="unclimbed",]
unclimbed
```

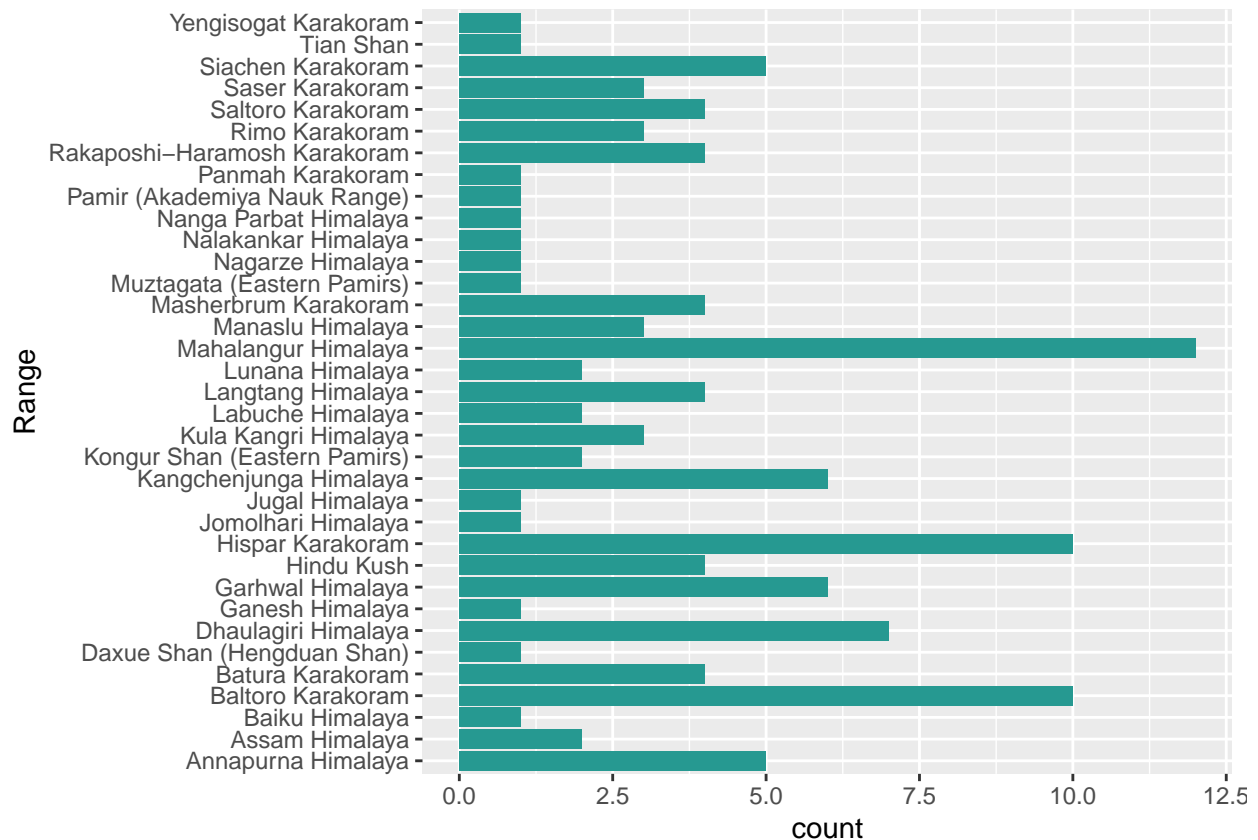
##	Rank	Mountain	Height (m)	Height (ft)	Prominence (m)
## 45	40	Gangkhar Puensum	7570	24836	2995
## 103	94	Labuche Kang III / East	7250	23786	570
## 109	100	Karjiang	7221	23691	880
## 112	103	Tongshanjiabu	7207	23645	1757

##	Range	Coordinates	Parent mountain
## 45	Kula Kangri Himalaya	28°02'50"N 90°27'19"E<U+FEFF>	Kangchenjunga
## 103	Labuche Himalaya	28°18'05"N 86°23'02"E<U+FEFF>	Labuche Kang
## 109	Kula Kangri Himalaya	28°15'27"N 90°38'49"E<U+FEFF>	Kula Kangri
## 112	Lunana Himalaya	28°11'12"N 89°57'27"E<U+FEFF>	Gangkar Puensum

##	First ascent	Ascents bef. 2004.	Failed attempts bef. 2004.
## 45	unclimbed	0	3
## 103	unclimbed	0	0
## 109	unclimbed	0	2
## 112	unclimbed	0	0

Grupiramo li planinske vrhove po planinskim lancima i prebrojimo li njihov broj u pojedinom planinskom lancu dobivamo histogram:

```
ggplot(data,aes(x=Range))+geom_bar(fill = "#269991")+coord_flip()
```



Iz histograma možemo vidjeti da se 12 najviših planinskih vrhova od 118 u dataframu nalazi u **Mahalangur Himalaya** planinskom lancu. Planinski lanci Baltoro Karakoram i Hispar Karakoram imaju 10 vrhova među najvišljim u svijetu.

Prvih 14 najviših planinskih vrhova su višji od 8000 metara. Planinski vrhovi su poredani po visini i vidimo da je Mount Everest najvišlja vrh na svijetu.

```
over8000 <- data[data$`Height (m)`>=8000,]
over8000
```

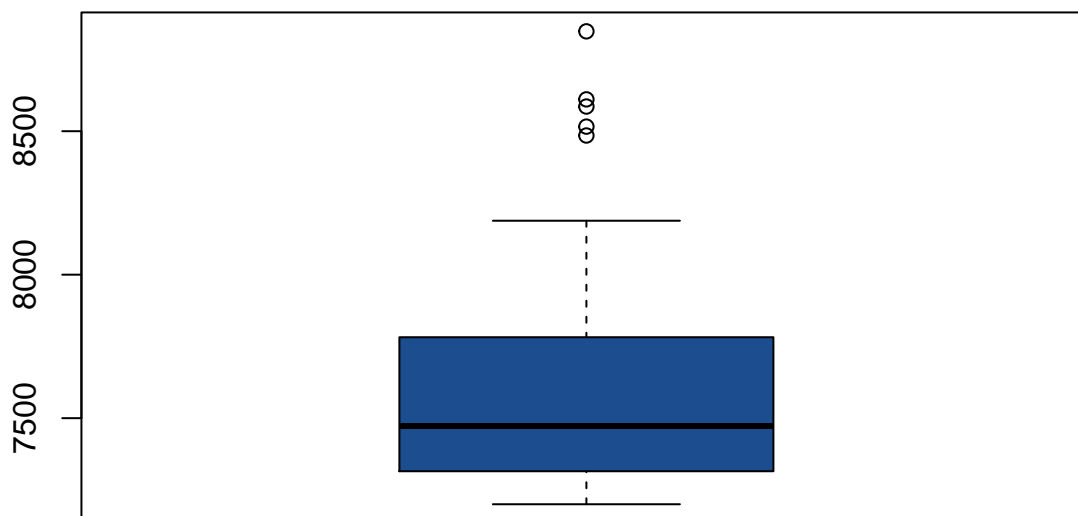
##	Rank	Mountain	Height (m)	Height (ft)
## 1	1	Mount Everest / Sagarmatha / Chomolungma	8848	29029
## 2	2	K2 / Qogir / Godwin Austen	8611	28251
## 3	3	Kangchenjunga	8586	28169
## 4	4	Lhotse	8516	27940
## 5	5	Makalu	8485	27838
## 6	6	Cho Oyu	8188	26864
## 7	7	Dhaulagiri I	8167	26795
## 8	8	Manaslu	8163	26781
## 9	9	Nanga Parbat	8126	26660
## 10	10	Annapurna I	8091	26545
## 11	11	Gasherbrum I / Hidden Peak / K5	8080	26509
## 12	12	Broad Peak / K3	8051	26414
## 13	13	Gasherbrum II / K4	8035	26362

## 14	14		Shishapangma	8027	26335
##	Prominence (m)		Range		Coordinates
## 1	8848	Mahalangur Himalaya	27°59'17"N	86°55'31"E	<U+FEFF>
## 2	4017	Baltoro Karakoram	35°52'53"N	76°30'48"E	<U+FEFF>
## 3	3922	Kangchenjunga Himalaya	27°42'12"N	88°08'51"E	<U+FEFF>
## 4	610	Mahalangur Himalaya	27°57'42"N	86°55'59"E	<U+FEFF>
## 5	2386	Mahalangur Himalaya	27°53'23"N	87°05'20"E	<U+FEFF>
## 6	2340	Mahalangur Himalaya	28°05'39"N	86°39'39"E	<U+FEFF>
## 7	3357	Dhaulagiri Himalaya	28°41'48"N	83°29'35"E	<U+FEFF>
## 8	3092	Manaslu Himalaya	28°33'00"N	84°33'35"E	<U+FEFF>
## 9	4608	Nanga Parbat Himalaya	35°14'14"N	74°35'21"E	<U+FEFF>
## 10	2984	Annapurna Himalaya	28°35'44"N	83°49'13"E	<U+FEFF>
## 11	2155	Baltoro Karakoram	35°43'28"N	76°41'47"E	<U+FEFF>
## 12	1701	Baltoro Karakoram	35°48'38"N	76°34'06"E	<U+FEFF>
## 13	1524	Baltoro Karakoram	35°45'28"N	76°39'12"E	<U+FEFF>
## 14	2897	Jugal Himalaya	28°21'12"N	85°46'43"E	<U+FEFF>
##	Parent mountain	First ascent	Ascents bef. 2004.	Failed attempts bef. 2004.	
## 1		1953	NA		121
## 2	Mount Everest	1954	45		44
## 3	Mount Everest	1955	38		24
## 4	Mount Everest	1956	26		26
## 5	Mount Everest	1955	45		52
## 6	Mount Everest	1954	79		28
## 7	K2	1960	51		39
## 8	Cho Oyu	1956	49		45
## 9	Dhaulagiri	1953	52		67
## 10	Cho Oyu	1950	36		47
## 11	K2	1958	31		16
## 12	Gasherbrum I	1957	39		19
## 13	Gasherbrum I	1956	54		12
## 14	Cho Oyu	1964	43		19

Analiza visine planinskih vrhova.

```
boxplot(data$`Height (m)`,
        names = c('Height in meters'),
        main='Boxplot : height in meters ',
        col="#1b4d8f")
```


Boxplot : height in meters



```
summary(data$`Height (m)`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      7200   7316   7472    7578   7776   8848
```

Srednja vrijednost, najveća i najmanja visina planinskih vrhova grupiranih po planinskim lancima.

```
data %>% group_by(Range) %>%
  summarize(mean_height = round(mean(`Height (m)`, na.rm = TRUE),2),
            max_height= max(`Height (m)`, na.rm = TRUE),
            min_height= min(`Height (m)`, na.rm = TRUE)) -> heightData
heightData
```

```
## # A tibble: 35 x 4
##   Range                mean_height max_height min_height
##   <chr>                <dbl>      <int>      <int>
## 1 Annapurna Himalaya    7651.        8091        7219
## 2 Assam Himalaya        7538        7782        7294
## 3 Baiku Himalaya        7281        7281        7281
## 4 Baltoro Karakoram     7820.        8611        7276
## 5 Batura Karakoram      7568.        7795        7388
## 6 Daxue Shan (Hengduan Shan) 7556        7556        7556
## 7 Dhaulagiri Himalaya   7585.        8167        7246
## 8 Ganesh Himalaya       7422        7422        7422
```

```
## 9 Garhwal Himalaya          7479.      7816      7242
## 10 Hindu Kush              7488      7708      7349
## # ... with 25 more rows
```

U dataframe se dodaje postotak neuspješnosti penjanja za pojedini vrh tako da se broj neuspješnih penjanja podijeli s brojem uspješnih i neuspješnih penjanja.

```
dataAnalyze$`Failed attempts (%)` <-
  round(dataAnalyze$`Failed attempts bef. 2004.`/
    (dataAnalyze$`Failed attempts bef. 2004.`+dataAnalyze$`Ascents bef. 2004.`)*100,2)
head(dataAnalyze)
```

```
## Rank Mountain Height (m) Height (ft) Prominence (m)
## 2 2 K2 / Qogir / Godwin Austen 8611 28251 4017
## 3 3 Kangchenjunga 8586 28169 3922
## 4 4 Lhotse 8516 27940 610
## 5 5 Makalu 8485 27838 2386
## 6 6 Cho Oyu 8188 26864 2340
## 7 7 Dhaulagiri I 8167 26795 3357
## Range Coordinates Parent mountain
## 2 Baltoro Karakoram 35°52'53"N 76°30'48"E<U+FEFF> Mount Everest
## 3 Kangchenjunga Himalaya 27°42'12"N 88°08'51"E<U+FEFF> Mount Everest
## 4 Mahalangur Himalaya 27°57'42"N 86°55'59"E<U+FEFF> Mount Everest
## 5 Mahalangur Himalaya 27°53'23"N 87°05'20"E<U+FEFF> Mount Everest
## 6 Mahalangur Himalaya 28°05'39"N 86°39'39"E<U+FEFF> Mount Everest
## 7 Dhaulagiri Himalaya 28°41'48"N 83°29'35"E<U+FEFF> K2
## First ascent Ascents bef. 2004. Failed attempts bef. 2004.
## 2 1954 45 44
## 3 1955 38 24
## 4 1956 26 26
## 5 1955 45 52
## 6 1954 79 28
## 7 1960 51 39
## Failed attempts (%)
## 2 49.44
## 3 38.71
## 4 50.00
## 5 53.61
## 6 26.17
## 7 43.33
```

Test o dvije proporcije

Usporedba planinskih lanaca : “Kangchenjunga Himalaya” i “Garhwal Himalaya”

Analiziraju se planinski lanci “Baltoro Karakoram” i “Hispar Karakoram” koji imaju jednak broj vrhova među najvišjim u svijetu.

H0: neuspjela penjanja su jednako ili više zastupljena u “Kangchenjunga Himalaya” nego “Kangchenjunga Himalaya”

H1: neuspjela penjanja su manje zastupljena u “Kangchenjunga Himalaya” nego “Kangchenjunga Himalaya”

```
KangchenjungaHimalaya <- dataAnalyze[dataAnalyze$Range=="Kangchenjunga Himalaya", ]
KangchenjungaHimalaya
```

##	Rank	Mountain	Height (m)	Height (ft)	Prominence (m)
## 3	3	Kangchenjunga	8586	28169	3922
## 34	32	Jannu	7711	25299	1036
## 62	57	Jongsong Peak	7462	24482	1298
## 71	65	Kabru N	7412	24318	780
## 82	76	Kirat Chuli	7362	24153	1168
## 85	117	Gimmigela / The Twins	7350	24114	432
##		Range		Coordinates	Parent mountain
## 3		Kangchenjunga Himalaya	27°42'12"N	88°08'51"E<U+FEFF>	Mount Everest
## 34		Kangchenjunga Himalaya	27°40'56"N	88°02'40"E<U+FEFF>	Kangchenjunga
## 62		Kangchenjunga Himalaya	27°52'54"N	88°08'09"E<U+FEFF>	Kangchenjunga
## 71		Kangchenjunga Himalaya	27°38'02"N	88°07'00"E<U+FEFF>	Kangchenjunga
## 82		Kangchenjunga Himalaya	27°47'16"N	88°11'43"E<U+FEFF>	Kangchenjunga
## 85		Kangchenjunga Himalaya	27°44'27"N	88°09'31"E<U+FEFF>	Kangchenjunga
##		First ascent	Ascents bef. 2004.	Failed attempts bef. 2004.	
## 3		1955	38	24	
## 34		1962	17	12	
## 62		1930	2	3	
## 71		1994	1	2	
## 82		1939	1	6	
## 85		1994	3	1	
##		Failed attempts (%)			
## 3		38.71			
## 34		41.38			
## 62		60.00			
## 71		66.67			
## 82		85.71			
## 85		25.00			

```
GarhwalHimalaya <- dataAnalyze[dataAnalyze$Range=="Garhwal Himalaya", ]
GarhwalHimalaya
```

##	Rank	Mountain	Height (m)	Height (ft)	Prominence (m)	Range	
## 25	23	Nanda Devi	7816	25643	3139	Garhwal Himalaya	
## 31	29	Kamet	7756	25446	2825	Garhwal Himalaya	
## 66	115	Sunanda Devi	7434	24390	260	Garhwal Himalaya	
## 83	116	Abi Gamin	7355	24131	217	Garhwal Himalaya	
## 100	92	Mana Peak	7272	23858	730	Garhwal Himalaya	
## 106	97	Mukut Parbat	7242	23760	840	Garhwal Himalaya	
##				Coordinates	Parent mountain	First ascent	
## 25				30°22'33"N	79°58'15"E<U+FEFF>	Dhaulagiri	1936
## 31				30°55'12"N	79°35'30"E<U+FEFF>	Nanda Devi	1931
## 66				30°22'00"N	79°59'40"E<U+FEFF>	Nanda Devi	1939
## 83				30°55'57"N	79°36'09"E<U+FEFF>	Kamet	1950
## 100				30°52'50"N	79°36'55"E<U+FEFF>	Kamet	1937
## 106				30°56'57"N	79°34'12"E<U+FEFF>	Kamet	1951
##				Ascents bef. 2004.	Failed attempts bef. 2004.	Failed attempts (%)	
## 25				14	12	46.15	
## 31				23	14	37.84	

## 66	14	12	46.15
## 83	17	2	10.53
## 100	7	3	30.00
## 106	2	1	33.33

Zbraja se ukupan broj neuspješnih i uspješnih penjanja za Kangchenjunga Himalaya i Garhwal Himalaya.

```
sumFailedKangchenjunga <-
  sum(dataAnalyze[dataAnalyze$Range=="Kangchenjunga Himalaya", ]$`Failed attempts bef. 2004.`)
sumFailedKangchenjunga
```

```
## [1] 48
```

```
sumSuccKangchenjunga <-
  sum(dataAnalyze[dataAnalyze$Range=="Kangchenjunga Himalaya", ]$`Ascents bef. 2004.`)
sumSuccKangchenjunga
```

```
## [1] 62
```

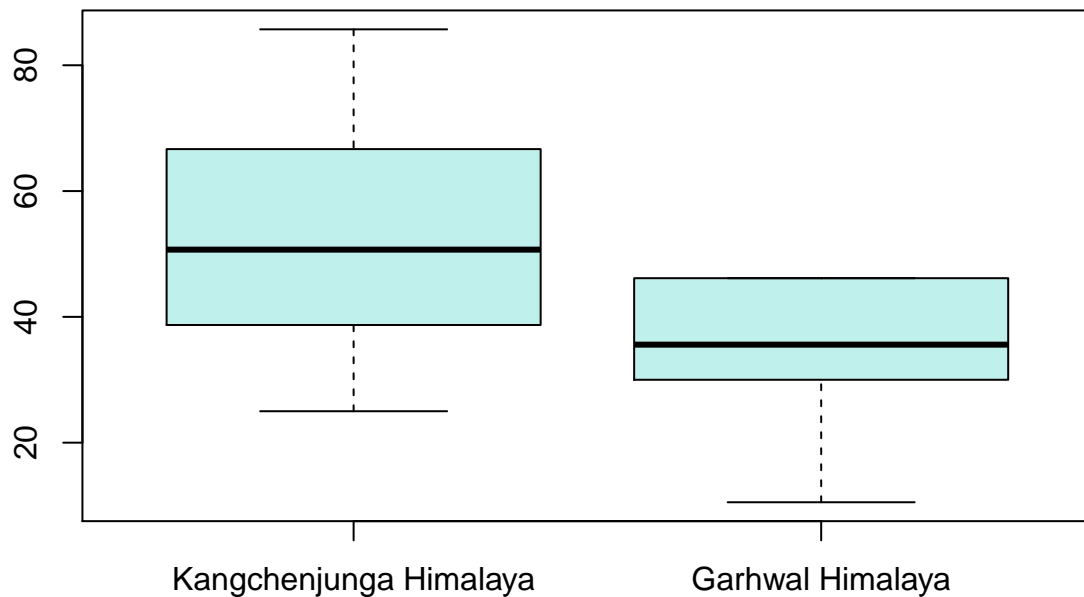
```
sumFailedGarhwal <-
  sum(dataAnalyze[dataAnalyze$Range=="Garhwal Himalaya", ]$`Failed attempts bef. 2004.`)
sumFailedGarhwal
```

```
## [1] 44
```

```
sumSuccGarhwal <-
  sum(dataAnalyze[dataAnalyze$Range=="Garhwal Himalaya", ]$`Ascents bef. 2004.`)
sumSuccGarhwal
```

```
## [1] 77
```

Boxplot of failed attempts in Kangchenjunga Himalaya and Garhwal Himalaya



Testiranje proporcija neuspješnosti penjanja.

```
failedAttempts <- c(sumFailedKangchenjunga,sumFailedGarhwal)
attempts <- c(sumFailedKangchenjunga+sumSuccKangchenjunga,sumSuccGarhwal+sumFailedGarhwal)

prop.test(failedAttempts, attempts,
          alternative = c("less"),
          conf.level = 0.95)
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data:  failedAttempts out of attempts
## X-squared = 0.98633, df = 1, p-value = 0.8397
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.0000000 0.1873462
## sample estimates:
##  prop 1    prop 2
## 0.4363636 0.3636364
```

Rezultat : P vrijednost provedenog testa iznosi 0.8397 uz 95% interval pouzdanosti.

p-vrijednost > **H0 se prihvća.**

Veću zastupljenost neuspjelih penjanja u “Kangchenjunga Himalaya” nego “Garhwal Himalaya” možemo iščitati iz usporedbe proporcija na testu gdje je ta razlika oko 7% veća u korist “Kangchenjunga Himalaya”.

Test homogenosti

```
sumFailedBaltoro <-  
  sum(dataAnalyze[dataAnalyze$Range=="Baltoro Karakoram", ]$`Failed attempts bef. 2004.`)  
sumFailedBaltoro
```

```
## [1] 110
```

```
sumSuccBaltoro <-  
  sum(dataAnalyze[dataAnalyze$Range=="Baltoro Karakoram", ]$`Ascents bef. 2004.`)  
sumSuccBaltoro
```

```
## [1] 183
```

```
sumFailedMahalangur <-  
  sum(dataAnalyze[dataAnalyze$Range=="Mahalangur Himalaya", ]$`Failed attempts bef. 2004.`)  
sumFailedMahalangur
```

```
## [1] 134
```

```
sumSuccMahalangur <-  
  sum(dataAnalyze[dataAnalyze$Range=="Mahalangur Himalaya", ]$`Ascents bef. 2004.`)  
sumSuccMahalangur
```

```
## [1] 185
```

```
sumFailedHispar <-  
  sum(dataAnalyze[dataAnalyze$Range=="Hispar Karakoram", ]$`Failed attempts bef. 2004.`)  
sumFailedHispar
```

```
## [1] 27
```

```
sumSuccHispar <-  
  sum(dataAnalyze[dataAnalyze$Range=="Hispar Karakoram", ]$`Ascents bef. 2004.`)  
sumSuccHispar
```

```
## [1] 19
```

Planinski lanci Mahalangur Himalaya, Baltoro Karakoram i Hispar Karakoram koji imaju najveći broj planinskih vrhova među najvišljim u svijetu imaju jednak postotak uspješnih i neuspješnih penjanja.

```
attempts <- data.frame(range =  
  c("Mahalangur Himalaya", "Baltoro Karakoram", "Hispar Karakoram", ""),  
  
  `successful attempts` = c(sumSuccMahalangur,  
                             sumSuccBaltoro,  
                             sumSuccHispar,  
                             sumSuccMahalangur+sumSuccBaltoro+sumSuccHispar),
```

```
`failed attempts` = c(sumFailedMahalangur,
                      sumFailedBaltoro,
                      sumFailedHispar,
                      sumFailedMahalangur+sumFailedBaltoro+sumFailedHispar),
`all attempts` = c(sumFailedMahalangur+sumSuccMahalangur,
                  sumFailedBaltoro+sumSuccBaltoro,
                  sumFailedHispar+sumSuccHispar,
                  sumFailedMahalangur+sumFailedBaltoro+sumFailedHispar+
                  sumSuccMahalangur+sumSuccBaltoro+sumSuccHispar))
```

```
attempts
```

```
##                range successful.attempts failed.attempts all.attempts
## 1 Mahalangur Himalaya                185                134                319
## 2 Baltoro Karakoram                  183                110                293
## 3 Hispar Karakoram                   19                 27                 46
## 4                                387                271                658
```

U kontingencijskoj tablici su prikazani broj uspješnih i neuspješnih penjanja za tri planinska lanca s najvećim brojem vrhova koji su u top 118 u svijetu. Provodi se testiranje hipoteze da su populacijske proporcije između svakog retka/stupca kontingencijske tablice jednake.

To je napravljeno testom homogenosti u kojem su hipoteze:

H0: postotak penjanja koji su uspješno završeni jednak je za sva tri planinska lanca,

H1: postotak penjanja koji su uspješno završeni nije jednak je za sva tri planinska lanca.

Provodi se Hi-kvadrat test nad podacima tablice.

$$\tilde{\chi}^2 = \frac{1}{d} \sum_{k=1}^n \frac{(O_k - E_k)^2}{E_k}$$

```
chisq <- chisq.test(attempts[1:3,2:3])
```

```
chisq
```

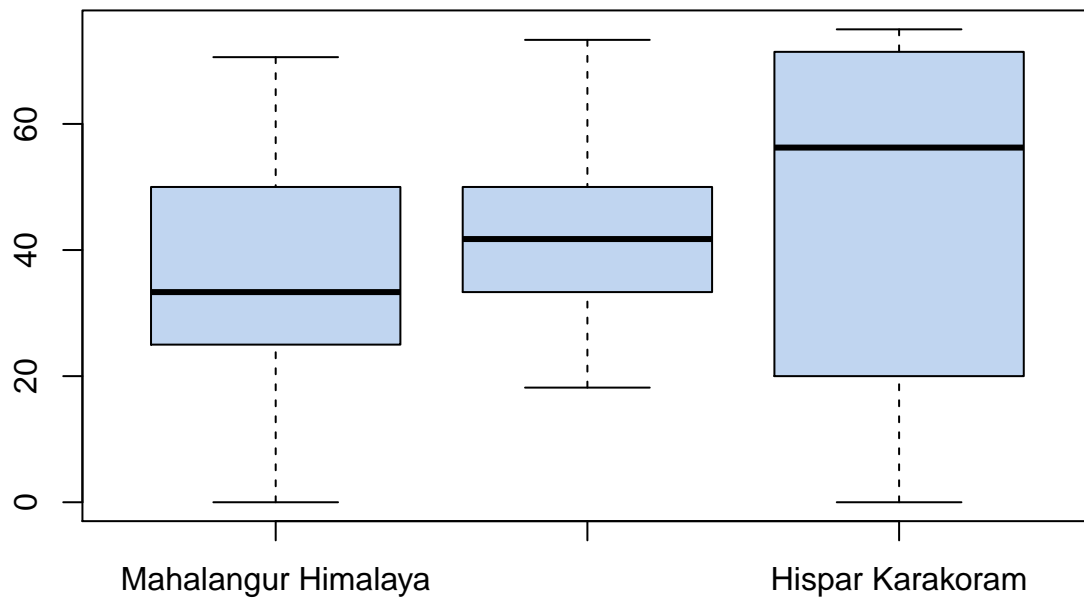
```
##
## Pearson's Chi-squared test
##
## data:  attempts[1:3, 2:3]
## X-squared = 7.5164, df = 2, p-value = 0.02333
```

Zaključak

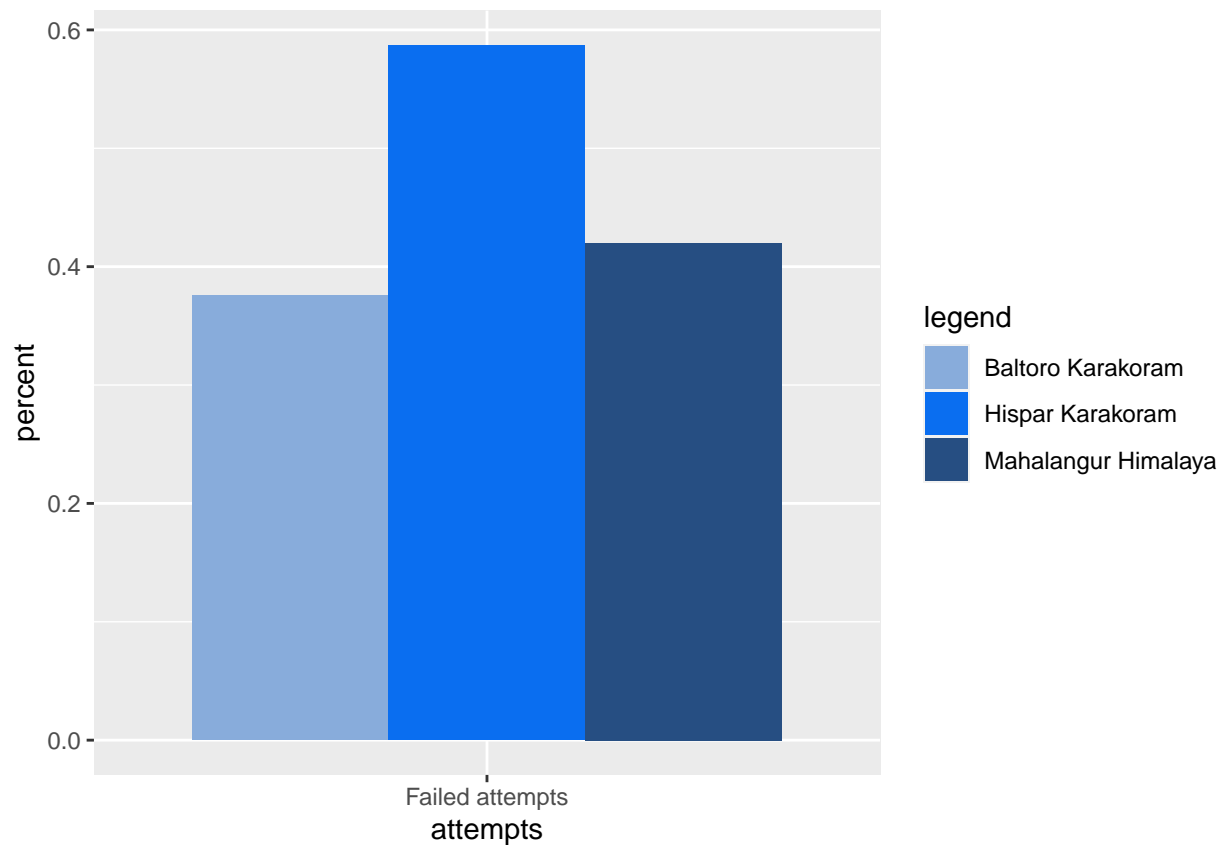
Provedbom testa homogenosti dobije se izuzetno mala p vrijednost, na temelju koje se može zaključiti da postotak uspješnih penjanja za Mahalangur Himalaya, Baltoro Karakoram i Hispar Karakoram nije jednak, te odbacujemo nultu hipotezu.

```
##          attempts  percent          range
## 1 Failed attempts 0.4200627 Mahalangur Himalaya
## 2 Failed attempts 0.3754266 Baltoro Karakoram
## 3 Failed attempts 0.5869565 Hispar Karakoram
```

Boxplot of failed attempts



```
ggplot(proportions) +  
  geom_col(  
    aes(x = attempts, y = percent, fill = range),  
    position = 'dodge') +  
  scale_fill_manual("legend", values = c("Mahalangur Himalaya" = "#264e82",  
                                          "Baltoro Karakoram" = "#88acdb",  
                                          "Hispar Karakoram" = "#0a6ef0"))
```

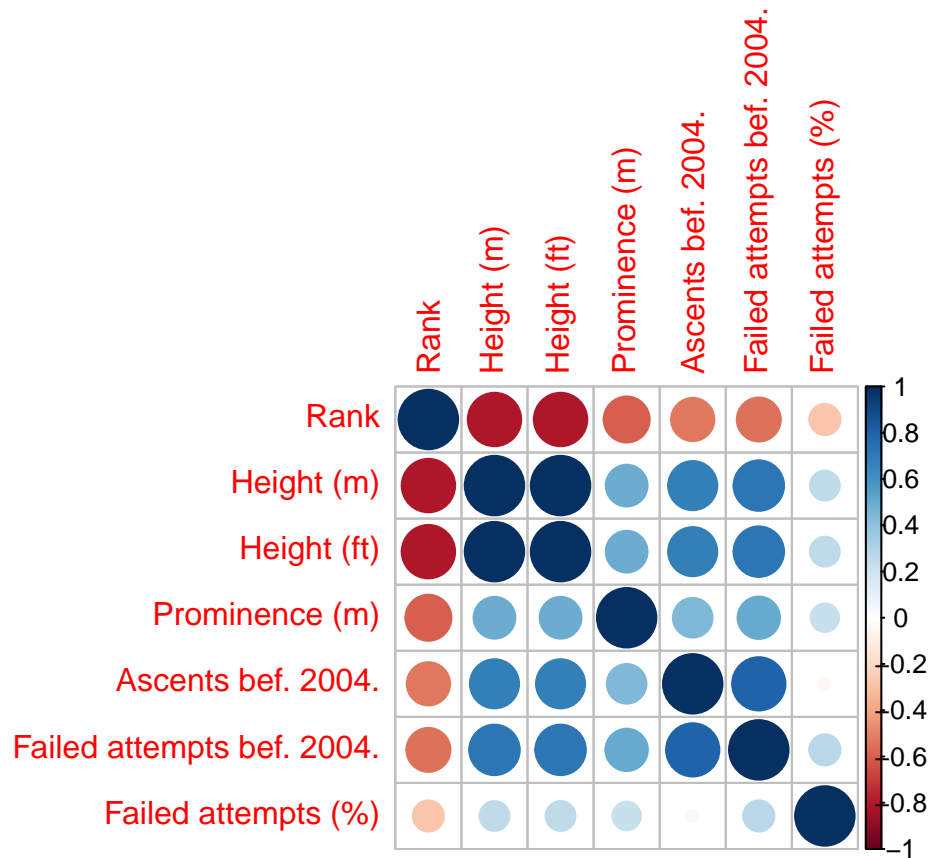



Boxplot i stupčasti dijagram potvrđuju provedbu testa. Uočava se da najveći postotak neuspješnih penjanja ima Hispar Karakoram.

Linearna regresija

Graf prikazuj međusobnu zavisnost varijabli. Veći i tamniji krugovi predstavljaju veću zavisnost, a manji i svjetliji manju.

```
numeric_col <- sapply(dataAnalyze, is.numeric)
df_numeric <- dataAnalyze[, numeric_col]
Correlation <- cor(df_numeric, use="pairwise.complete.obs")
corrplot(Correlation)
```



Provodi se linearna regresija s obzirom na prikaz grafa tako da se odabiru nezavisni regresori.

```
##
## Call:
## lm(formula = `Ascents bef. 2004.` ~ `Height (m)`, data = dataAnalyze)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.561  -6.686  -0.217   3.533  51.038
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.225e+02  2.342e+01  -9.498 4.77e-16 ***
## `Height (m)`  3.058e-02  3.091e-03   9.893 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.71 on 112 degrees of freedom
## Multiple R-squared:  0.4663, Adjusted R-squared:  0.4616
## F-statistic: 97.87 on 1 and 112 DF, p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = `Ascents bef. 2004.` ~ `Height (m)` + `Prominence (m)`,
```

```

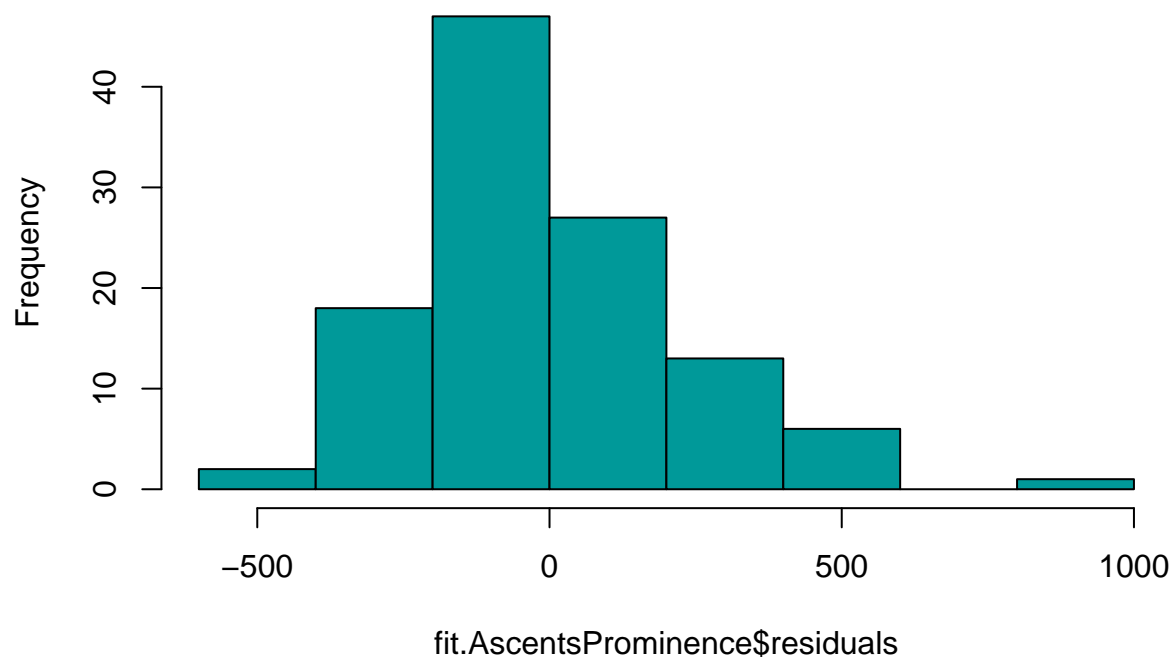
##      data = dataAnalyze)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -19.110   -6.306   -0.867    3.750   51.384
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.019e+02  2.581e+01  -7.820 3.30e-12 ***
## `Height (m)`    2.744e-02  3.517e-03   7.801 3.64e-12 ***
## `Prominence (m)` 2.058e-03  1.134e-03   1.814  0.0724 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.6 on 111 degrees of freedom
## Multiple R-squared:  0.4817, Adjusted R-squared:  0.4723
## F-statistic: 51.58 on 2 and 111 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = `Height (m)` ~ `Ascents bef. 2004.` + `Prominence (m)`,
##     data = dataAnalyze)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -421.56  -165.96   -38.05   113.59   800.43
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.334e+03  3.981e+01 184.211 < 2e-16 ***
## `Ascents bef. 2004.` 1.291e+01  1.654e+00   7.801 3.64e-12 ***
## `Prominence (m)`    7.613e-02  2.390e-02   3.185  0.00188 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 229.9 on 111 degrees of freedom
## Multiple R-squared:  0.511, Adjusted R-squared:  0.5022
## F-statistic:   58 on 2 and 111 DF,  p-value: < 2.2e-16

hist(fit.AscentsProminence$residuals, col="#009999")

```

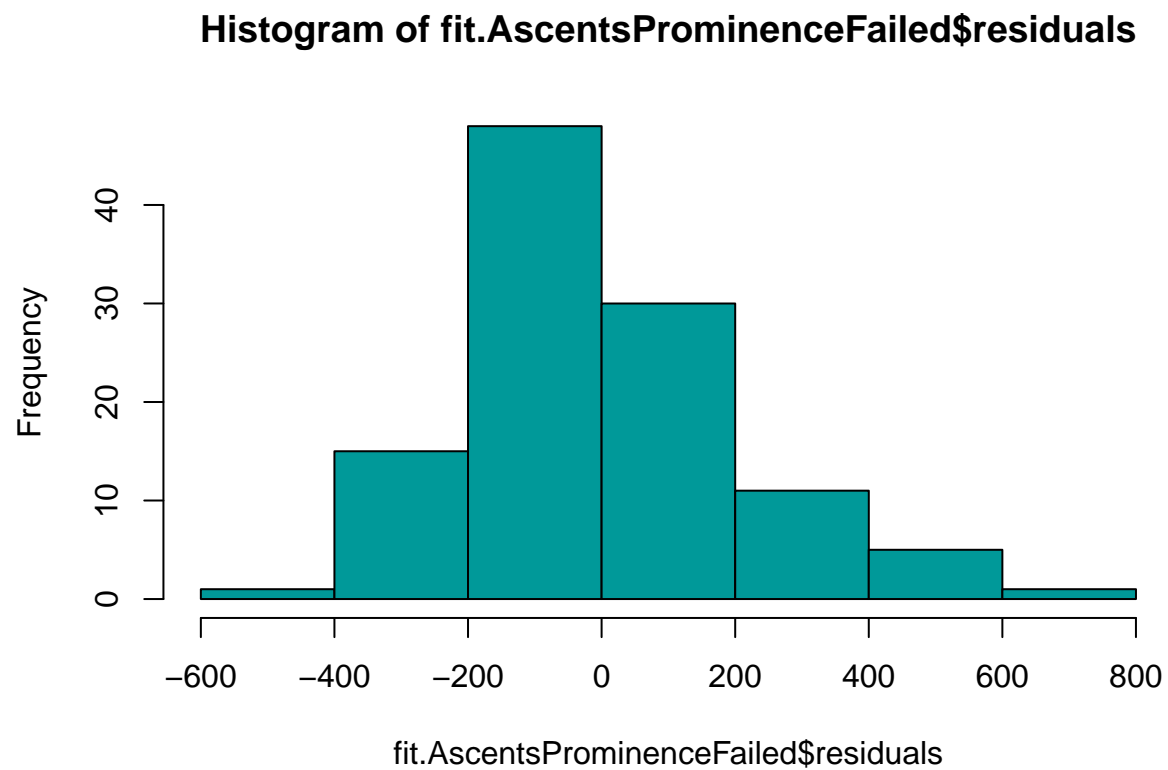
Histogram of fit.AscentsProminence\$residuals



```
##
## Call:
## lm(formula = `Height (m)` ~ `Ascents bef. 2004.` + `Prominence (m)` +
##     `Failed attempts (%)`, data = dataAnalyze)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -431.9  -127.8   -52.6   108.3   735.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.251e+03  4.540e+01 159.723 < 2e-16 ***
## `Ascents bef. 2004.` 1.370e+01  1.609e+00  8.512 1.15e-13 ***
## `Prominence (m)`    5.289e-02  2.385e-02  2.217 0.028724 *
## `Failed attempts (%)` 2.811e+00  7.880e-01  3.567 0.000542 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 219.6 on 107 degrees of freedom
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.5609, Adjusted R-squared:  0.5485
## F-statistic: 45.55 on 3 and 107 DF, p-value: < 2.2e-16
```

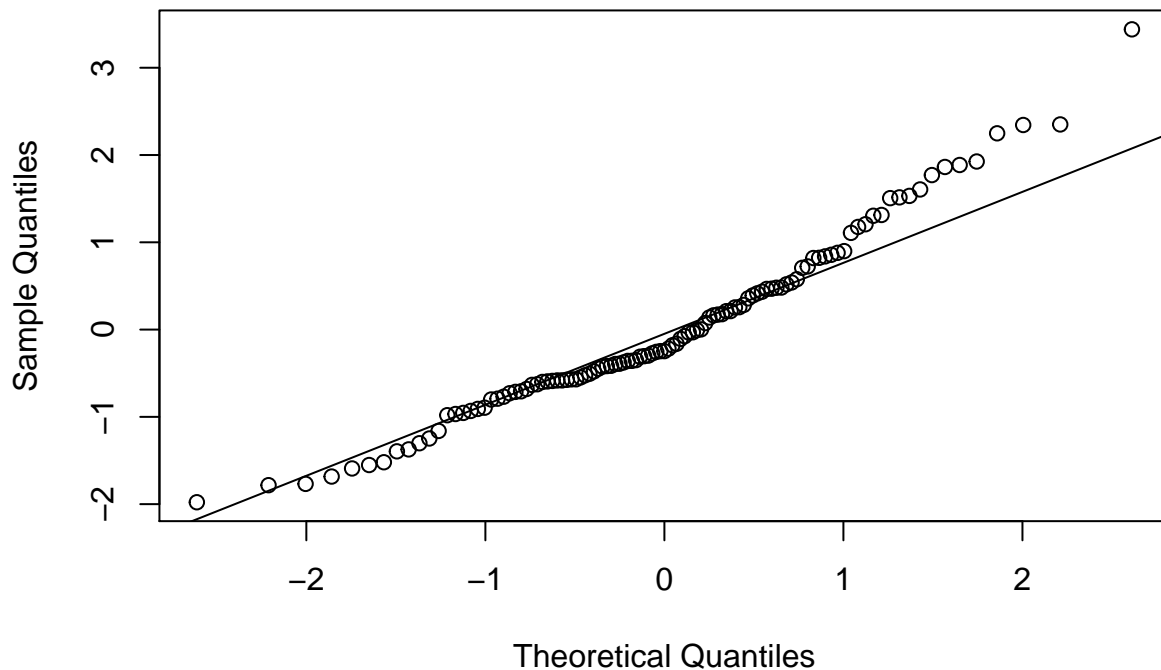
Normalnost reziduala

```
hist(fit.AscentsProminenceFailed$residuals, col="#009999")
```



```
qqnorm(rstandard(fit.AscentsProminenceFailed))  
qqline(rstandard(fit.AscentsProminenceFailed))
```

Normal Q-Q Plot



```
##
## Call:
## lm(formula = `Failed attempts (%)` ~ `Prominence (m)` + `Height (m)`,
##     data = dataAnalyze)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41.535 -22.546  -2.816   21.721   67.407
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -82.754989   66.027018  -1.253   0.2128
## `Prominence (m)`    0.003788    0.002883   1.314   0.1917
## `Height (m)`       0.015512    0.008988   1.726   0.0872 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.81 on 108 degrees of freedom
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.07827,    Adjusted R-squared:  0.0612
## F-statistic: 4.585 on 2 and 108 DF,  p-value: 0.01226
```

Od svih provedenih modela najveći korelacijski koeficijent je dobiven u modelu gdje su broj uspješnih i neuspješnih penjanja i prominencija regresori za nadmorsku visinu planinskog vrha. Histogramom i qqplotom ispitana je normalnost reziduala tog modela i zaključuje se da je on valjani model.