WS3 Assignment of Data Analyses

Jari Haukka

2021-08-30

Contents

Assignment

1. Read carefully data analyses below (tables, figures, regression analyses)
2. Write description of data using tables and figures. In this assignment we use hospital stay lenght as key performance indicator (KPI). You may use copy-paste to include tables and figures in you report.
3. Comment briefly raw data. Is there any interesting features?
4. Interpret both regression regression models below. Did adding variable `Payment Typology 1` change interpretation of model?

Read open data

Data read from NYSDOH_HospitalInpatientDischarges_SPARCS_De-Identified_2012.csv.

Original Title: Hospital Inpatient Discharges (SPARCS De-Identified) Downloadable File: 2012

The Statewide Planning and Research Cooperative System (SPARCS) Inpatient De-identified downloadable file contains discharge level detail on patient characteristics, diagnoses, treatments, services, and charges. This data file contains basic record level detail for the discharge. The de-identified data file does not contain data that is protected health information (PHI) under HIPAA. The health information is not individually identifiable; all data elements considered identifiable have been redacted. For example, the direct identifiers regarding a date have the day and month portion of the date removed. For more information, including changes to the data from previous years, please visit http://www.health.ny.gov/statistics/sparcs/access/.

Source: https://health.data.ny.gov/d/3m9u-ws8e Last updated at https://health.data.ny.gov : 2019-09-13

```r
# Load packages
library(xtable)
library(Epi)
library(foreign)
library(survival)
library(epitools)
library(tableone)
library(readr)

# eHeaTZ.NYSDOH <- as.data.frame(
#   read_csv("NYSDOH_HospitalInpatientDischarges_SPARCS_De-Identified_2012.csv"))
# Load R data
# load(file="eHeaTZ.NYSDOH.RData")
# Use only New York City data
# eHeaTZ.NYSDOH.1<-subset(eHeaTZ.NYSDOH,`Hospital Service Area`=="New York City")#5
# eHeaTZ.NYSDOH.1<-subset(eHeaTZ.NYSDOH,`Hospital Service Area`=="Southern Tier")#2
# save(eHeaTZ.NYSDOH.1,file="eHeaTZ.NYSDOH.1.RData")

# eHeaTZ.NYSDOH.1<-subset(eHeaTZ.NYSDOH,`Facility Name`=="North Shore University Hospital")#2
# save(eHeaTZ.NYSDOH.1,file="eHeaTZ.NYSDOH.1A.RData")

# head(rev(sort(table(eHeaTZ.NYSDOH$`Facility Name`))))
load(file="eHeaTZ.NYSDOH.1A.RData")
```

Checking data with simple summaries

We use "North Shore University Hospital" data in this example

```r
# First fivee rows of the data
head(eHeaTZ.NYSDOH.1,5)
```

```
##         Hospital Service Area Hospital County Operating Certificate Number
## 2320769           Long Island          Nassau                      2951001
## 2320770           Long Island          Nassau                      2951001
## 2320771           Long Island          Nassau                      2951001
## 2320772           Long Island          Nassau                      2951001
## 2320773           Long Island          Nassau                      2951001
##         Permanent Facility Id                   Facility Name Age Group
## 2320769                000541 North Shore University Hospital  18 to 29
## 2320770                000541 North Shore University Hospital  18 to 29
## 2320771                000541 North Shore University Hospital   0 to 17
## 2320772                000541 North Shore University Hospital   0 to 17
## 2320773                000541 North Shore University Hospital   0 to 17
##         Zip Code - 3 digits Gender                    Race         Ethnicity
## 2320769                 110      F                   White Not Span/Hispanic
## 2320770                 113      M Black/African American Not Span/Hispanic
## 2320771                 113      M                   White Not Span/Hispanic
## 2320772                 113      M                   White Not Span/Hispanic
## 2320773                 104      M                   White  Spanish/Hispanic
##         Length of Stay Type of Admission                  Patient Disposition
## 2320769             NA         Emergency Psychiatric Hospital or Unit of Hosp
## 2320770             NA         Emergency Psychiatric Hospital or Unit of Hosp
## 2320771             NA           Newborn         Home w/ Home Health Services
```

```
## 2320772            117           Newborn         Home w/ Home Health Services
## 2320773             NA           Newborn                   Short-term Hospital
##         Discharge Year CCS Diagnosis Code
## 2320769           2012                657
## 2320770           2012                659
## 2320771           2012                218
## 2320772           2012                218
## 2320773           2012                218
##                          CCS Diagnosis Description CCS Procedure Code
## 2320769                             Mood disorders                218
## 2320770 Schizophrenia and other psychotic disorders                218
## 2320771                                  Liveborn                061
## 2320772                                  Liveborn                002
## 2320773                                  Liveborn                216
##         CCS Procedure Description APR DRG Code
## 2320769   PSYCHO/PSYCHI EVAL/THER          753
## 2320770   PSYCHO/PSYCHI EVAL/THER          750
## 2320771   OT OR PRCS VES NOT HEAD          588
## 2320772   IRR XCRANIAL VENT SHUNT          588
## 2320773    RESP INTUB/MECH VENTIL          593
##                                APR DRG Description APR MDC Code
## 2320769                          Bipolar disorders           19
## 2320770                              Schizophrenia           19
## 2320771         Neonate bwt <1500g w major procedure           15
## 2320772         Neonate bwt <1500g w major procedure           15
## 2320773 Neonate birthwt 750-999g w/o major procedure           15
##                                                                   APR MDC Description
## 2320769                                                   Mental Diseases and Disorders
## 2320770                                                   Mental Diseases and Disorders
## 2320771 Newborns and Other Neonates with Conditions Originating in the Perinatal Period
## 2320772 Newborns and Other Neonates with Conditions Originating in the Perinatal Period
## 2320773 Newborns and Other Neonates with Conditions Originating in the Perinatal Period
##         APR Severity of Illness Code APR Severity of Illness Description
## 2320769                            2                            Moderate
## 2320770                            2                            Moderate
## 2320771                            4                             Extreme
## 2320772                            4                             Extreme
## 2320773                            4                             Extreme
##         APR Risk of Mortality APR Medical Surgical Description
## 2320769                 Minor                          Medical
## 2320770                 Minor                          Medical
## 2320771                 Major                         Surgical
## 2320772                 Major                         Surgical
## 2320773                 Major                          Medical
##               Payment Typology 1       Payment Typology 2
## 2320769                 Medicare Private Health Insurance
## 2320770 Private Health Insurance                 Medicaid
## 2320771   Blue Cross/Blue Shield                 Medicaid
## 2320772   Blue Cross/Blue Shield                 Medicaid
## 2320773                 Medicaid                     <NA>
##               Payment Typology 3 Birth Weight Abortion Edit Indicator
## 2320769                 Medicaid         0000                       N
## 2320770 Private Health Insurance         0000                       N
## 2320771                     <NA>         0700                       N
```

```
## 2320772                    <NA>         0700                       N
## 2320773                    <NA>         0700                       N
##         Emergency Department Indicator Total Charges Total Costs
## 2320769                              Y      649018.5    134445.8
## 2320770                              Y      561857.0    116108.0
## 2320771                              N     1606846.3    264550.3
## 2320772                              N     1288911.5    207560.5
## 2320773                              N     1406583.0    225036.0
##         Ratio of Total Costs to Total Charges
## 2320769                             0.2071526
## 2320770                             0.2066505
## 2320771                             0.1646394
## 2320772                             0.1610355
## 2320773                             0.1599877
```

Descriptive tables / `onetable`-package

```r
# Tabulations, more complex, add more variables
# dput(names(eHeaTZ.NYSDOH))

# with(subset(eHeaTZ.NYSDOH,`Hospital Service Area`=="Central NY"),
#        table(`Hospital Service Area`, `Hospital County`))


#
tmp.vars<-c("Age Group","Gender","Payment Typology 1", "Emergency Department Indicator",
            "Length of Stay","Type of Admission","APR Severity of Illness Description",
            "Total Charges", "Total Costs","Ratio of Total Costs to Total Charges")

tmp.fvars<-c("Age Group","Gender","Type of Admission","APR Severity of Illness Description",
             "Payment Typology 1", "Emergency Department Indicator")

tmp.tableOne1 <- CreateTableOne(vars = tmp.vars ,factorVars=tmp.fvars,
                                strata = c("Gender"),
                                data = eHeaTZ.NYSDOH.1)

tmp.tableOne2 <- CreateTableOne(vars = tmp.vars ,factorVars=tmp.fvars,
                                strata = c("Type of Admission"),
                                data = eHeaTZ.NYSDOH.1)

tmp.tableOne3 <- CreateTableOne(vars = tmp.vars ,factorVars=tmp.fvars,
                                strata = c("Payment Typology 1"),
                                data = eHeaTZ.NYSDOH.1)
```

```r
# Tabulations, more complex, add more variables
# Print results
knitr::kable(print(tmp.tableOne1,missing=TRUE,printToggle = FALSE),
             caption="Tables by gender")
```

Table 1: Tables by gender

	F	M	p	test	Missing
n	30265	21603			
Age Group (%)			<0.001		0.0
0 to 17	3625 (12.0)	3838 (17.8)			
18 to 29	2996 (9.9)	849 (3.9)			
30 to 49	8189 (27.1)	2711 (12.5)			
50 to 69	6382 (21.1)	6676 (30.9)			
70 or Older	9073 (30.0)	7529 (34.9)			
Gender = M (%)	0 (0.0)	21603 (100.0)	<0.001		0.0
Payment Typology 1 (%)			<0.001		0.0
Blue Cross/Blue Shield	6342 (21.0)	3826 (17.7)			
Federal/State/Local/VA	1 (0.0)	4 (0.0)			
Medicaid	1024 (3.4)	910 (4.2)			
Medicare	11166 (36.9)	9538 (44.2)			
Private Health Insurance	11395 (37.7)	6863 (31.8)			
Self-Pay	101 (0.3)	89 (0.4)			
Unknown	236 (0.8)	373 (1.7)			
Emergency Department Indicator = Y (%)	14938 (49.4)	13009 (60.2)	<0.001		0.0
Length of Stay (mean (SD))	4.90 (6.04)	5.67 (7.67)	<0.001		0.0
Type of Admission (%)			<0.001		0.0
Elective	8463 (28.0)	2750 (12.7)			
Emergency	16797 (55.5)	13217 (61.2)			
Newborn	3119 (10.3)	3189 (14.8)			
Not Available	2 (0.0)	1 (0.0)			
Urgent	1884 (6.2)	2446 (11.3)			
APR Severity of Illness Description (%)			<0.001		0.0
Extreme	1698 (5.6)	1949 (9.0)			
Major	5632 (18.6)	4885 (22.6)			
Minor	12827 (42.4)	7667 (35.5)			
Moderate	10107 (33.4)	7102 (32.9)			
Total Charges (mean (SD))	52754.53 (71436.05)	65778.40 (94484.07)	<0.001		0.0
Total Costs (mean (SD))	12028.38 (16946.54)	14849.67 (22482.63)	<0.001		0.0
Ratio of Total Costs to Total Charges (mean (SD))	0.23 (0.04)	0.21 (0.04)	<0.001		0.0

```
knitr::kable(print(tmp.tableOne2,missing=TRUE,printToggle = FALSE),
             caption="Tables by type of admission")
```

Table 2: Tables by type of admission

	Elective	Emergency	Newborn	Not Available	Urgent	p	test	Missing
n	11213	30014	6308	3	4330			
Age Group (%)						<0.001		0.0
0 to 17	116 (1.0)	693 (2.3)	6308 (100.0)	1 (33.3)	345 (8.0)			
18 to 29	1685 (15.0)	1953 (6.5)	0 (0.0)	0 (0.0)	207 (4.8)			

	Elective	Emergency	Newborn	Not Available	Urgent	p test	Missing
30 to 49	4654 (41.5)	5575 (18.6)	0 (0.0)	0 (0.0)	671 (15.5)		
50 to 69	2763 (24.6)	8728 (29.1)	0 (0.0)	1 (33.3)	1566 (36.2)		
70 or Older	1995 (17.8)	13065 (43.5)	0 (0.0)	1 (33.3)	1541 (35.6)		
Gender = M (%)	2750 (24.5)	13217 (44.0)	3189 (50.6)	1 (33.3)	2446 (56.5)	<0.001	0.0
Payment Typology 1 (%)						<0.001	0.0
Blue Cross/Blue Shield	2033 (18.1)	5232 (17.4)	2277 (36.1)	0 (0.0)	626 (14.5)		
Federal/State/Local/VA	0 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	3 (0.1)		
Medicaid	277 (2.5)	1182 (3.9)	233 (3.7)	0 (0.0)	242 (5.6)		
Medicare	2773 (24.7)	15893 (53.0)	0 (0.0)	2 (66.7)	2036 (47.0)		
Private Health Insurance	5995 (53.5)	7132 (23.8)	3787 (60.0)	1 (33.3)	1343 (31.0)		
Self-Pay	37 (0.3)	121 (0.4)	11 (0.2)	0 (0.0)	21 (0.5)		
Unknown	98 (0.9)	452 (1.5)	0 (0.0)	0 (0.0)	59 (1.4)		
Emergency Department Indicator = Y (%)	10 (0.1)	27926 (93.0)	1 (0.0)	2 (66.7)	8 (0.2)	<0.001	0.0
Length of Stay (mean (SD))	3.79 (4.75)	5.92 (6.83)	3.57 (6.24)	2.00 (1.73)	6.53 (9.85)	<0.001	0.0
Type of Admission (%)						<0.001	0.0
Elective	11213 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		
Emergency	0 (0.0)	30014 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)		
Newborn	0 (0.0)	0 (0.0)	6308 (100.0)	0 (0.0)	0 (0.0)		
Not Available	0 (0.0)	0 (0.0)	0 (0.0)	3 (100.0)	0 (0.0)		
Urgent	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4330 (100.0)		
APR Severity of Illness Description (%)						<0.001	0.0
Extreme	186 (1.7)	2919 (9.7)	46 (0.7)	0 (0.0)	496 (11.5)		
Major	1075 (9.6)	8163 (27.2)	289 (4.6)	1 (33.3)	989 (22.8)		
Minor	5972 (53.3)	7749 (25.8)	5351 (84.8)	2 (66.7)	1420 (32.8)		
Moderate	3980 (35.5)	11183 (37.3)	621 (9.8)	0 (0.0)	1425 (32.9)		
Total Charges (mean (SD))	54592.33 (71044.70)	62236.46 (75437.06)	24085.92 (81669.17)	24895.25 (16077.82)	89032.14 (125216.09)	<0.001	0.0
Total Costs (mean (SD))	14780.09 (19403.13)	13445.63 (17990.43)	3985.03 (13243.67)	5699.51 (3748.33)	20876.52 (29817.47)	<0.001	0.0
Ratio of Total Costs to Total Charges (mean (SD))	0.27 (0.03)	0.21 (0.03)	0.17 (0.01)	0.22 (0.04)	0.23 (0.04)	<0.001	0.0

```
knitr::kable(print(tmp.tableOne3,missing=TRUE,printToggle = FALSE),
             caption="Tables by Payment Typology 1" )
```

Table 3: Tables by Payment Typology 1

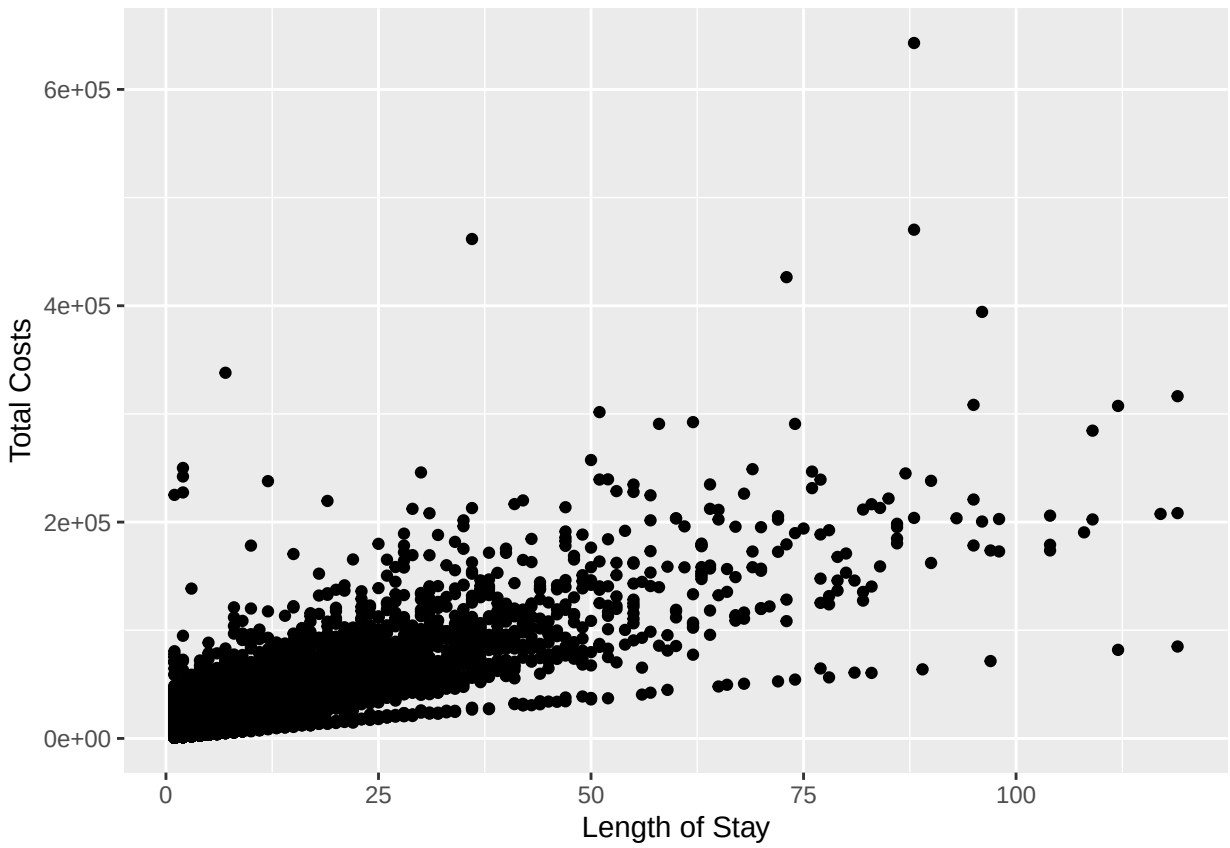
	Blue Cross/Blue Shield	Federal/State/Local/VA	Medicaid	Medicare	Private Health Insurance	Self-Pay	Unknown	p test	Missing
n	10168	5	1934	20704	18258	190	609		
Age Group (%)								<0.001	0.0
0 to 17	2560 (25.2)	0 (0.0)	372 (19.2)	0 (0.0)	4505 (24.7)	15 (7.9)	11 (1.8)		
18 to 29	961 (9.5)	0 (0.0)	295 (15.3)	136 (0.7)	2330 (12.8)	28 (14.7)	95 (15.6)		
30 to 49	3444 (33.9)	1 (20.0)	591 (30.6)	660 (3.2)	5970 (32.7)	61 (32.1)	173 (28.4)		
50 to 69	2940 (28.9)	4 (80.0)	528 (27.3)	4287 (20.7)	4978 (27.3)	84 (44.2)	237 (38.9)		
70 or Older	263 (2.6)	0 (0.0)	148 (7.7)	15621 (75.4)	475 (2.6)	2 (1.1)	93 (15.3)		
Gender = M (%)	3826 (37.6)	4 (80.0)	910 (47.1)	9538 (46.1)	6863 (37.6)	89 (46.8)	373 (61.2)	<0.001	0.0
Payment Typology 1 (%)								<0.001	0.0
Blue Cross/Blue Shield	10168 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		
Federal/State/Local/VA	0 (0.0)	5 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		
Medicaid	0 (0.0)	0 (0.0)	1934 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		
Medicare	0 (0.0)	0 (0.0)	0 (0.0)	20704 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)		
Private Health Insurance	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	18258 (100.0)	0 (0.0)	0 (0.0)		
Self-Pay	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	190 (100.0)	0 (0.0)		
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	609 (100.0)		
Emergency Department Indicator = Y (%)	3569 (35.1)	2 (40.0)	1136 (58.7)	15702 (75.8)	6975 (38.2)	121 (63.7)	442 (72.6)	<0.001	0.0
Length of Stay (mean (SD))	3.97 (5.39)	7.80 (4.60)	7.71 (11.98)	6.49 (7.09)	4.22 (6.00)	3.68 (4.75)	5.59 (6.40)	<0.001	0.0
Type of Admission (%)								<0.001	0.0
Elective	2033 (20.0)	0 (0.0)	277 (14.3)	2773 (13.4)	5995 (32.8)	37 (19.5)	98 (16.1)		
Emergency	5232 (51.5)	2 (40.0)	1182 (61.1)	15893 (76.8)	7132 (39.1)	121 (63.7)	452 (74.2)		
Newborn	2277 (22.4)	0 (0.0)	233 (12.0)	0 (0.0)	3787 (20.7)	11 (5.8)	0 (0.0)		
Not Available	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)		

	Blue Cross/Blue Shield	Federal/State/Local/MA	Medicaid	Medicare	Private Health Insurance	Self-Pay	Unknown	p	test	Missing
Urgent	626 (6.2)	3 (60.0)	242 (12.5)	2036 (9.8)	1343 (7.4)	21 (11.1)	59 (9.7)			
APR Severity of Illness Description (%)								<0.001		0.0
Extreme	323 (3.2)	0 (0.0)	180 (9.3)	2513 (12.1)	583 (3.2)	9 (4.7)	39 (6.4)			
Major	1113 (10.9)	4 (80.0)	372 (19.2)	6753 (32.6)	2154 (11.8)	16 (8.4)	105 (17.2)			
Minor	5795 (57.0)	0 (0.0)	708 (36.6)	3663 (17.7)	9976 (54.6)	102 (53.7)	250 (41.1)			
Moderate	2937 (28.9)	1 (20.0)	674 (34.9)	7775 (37.6)	5544 (30.4)	63 (33.2)	215 (35.3)			
Total Charges (mean (SD))	44381.65 (72860.99)	106641.82 (49202.86)	81521.98 (140074.58)	73387.38 (84473.55)	45705.01 (71987.87)	47511.39 (67279.82)	74279.30 (75913.63)	<0.001		0.0
Total Costs (mean (SD))	10248.62 (16490.77)	24784.77 (13809.05)	17804.33 (32596.26)	16511.27 (20690.37)	10477.55 (16946.47)	10859.57 (16271.39)	17830.67 (19398.00)	<0.001		0.0
Ratio of Total Costs to Total Charges (mean (SD))	0.22 (0.05)	0.22 (0.03)	0.21 (0.04)	0.22 (0.04)	0.22 (0.05)	0.21 (0.04)	0.22 (0.05)	<0.001		0.0

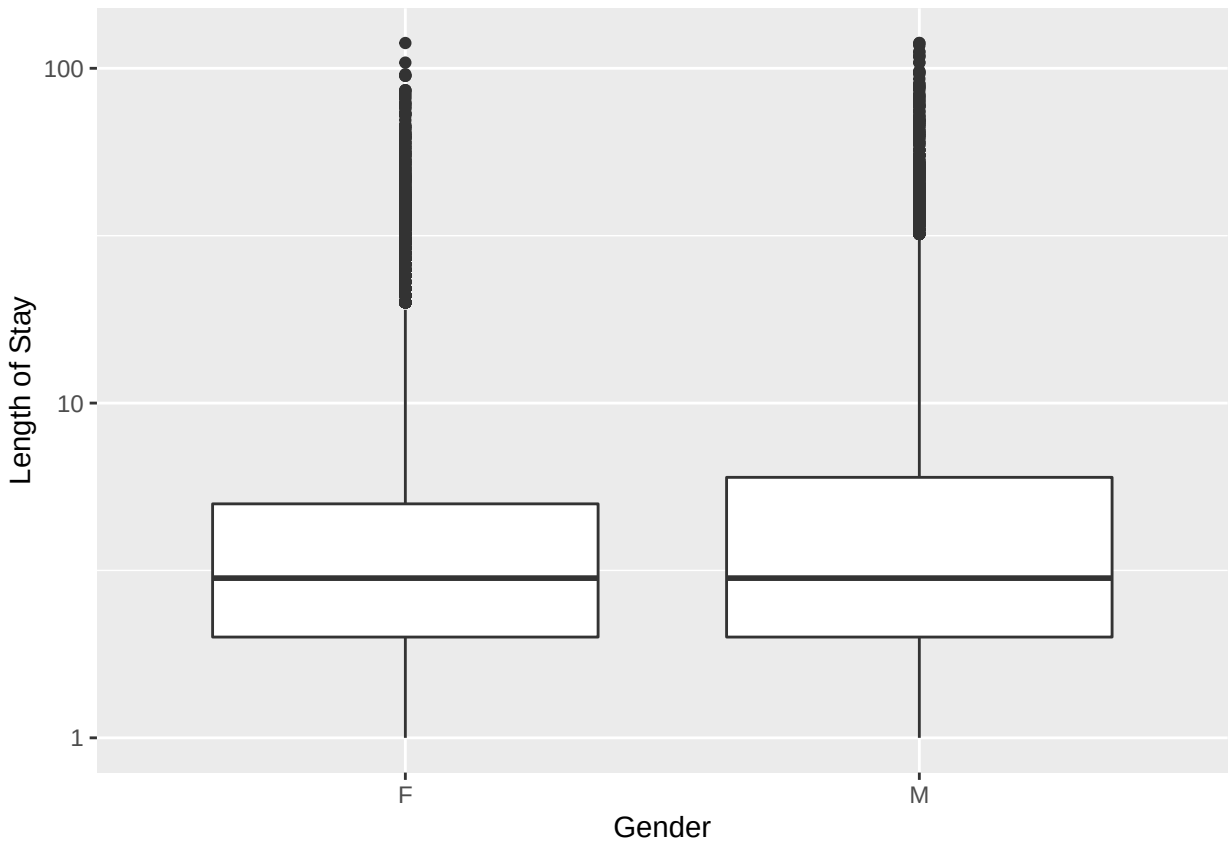
Visualization

Plot types

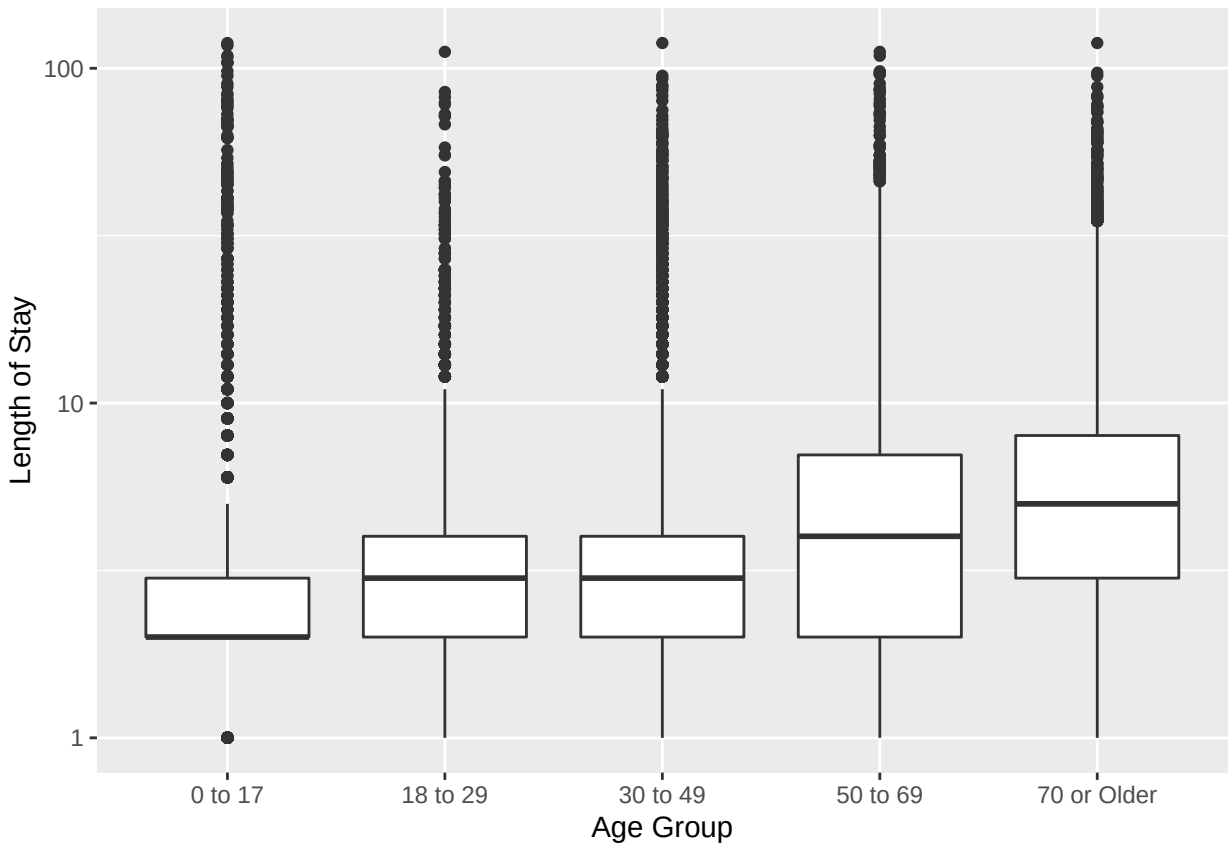
```
library(ggplot2)
library(plotly)
library(hrbrthemes)



tmp.p1<-ggplot(eHeaTZ.NYSDOH.1, aes(x=`Length of Stay`, y=`Total Costs`)) +
    geom_point() +  geom_smooth(method = "loess")
 (tmp.p1)
```



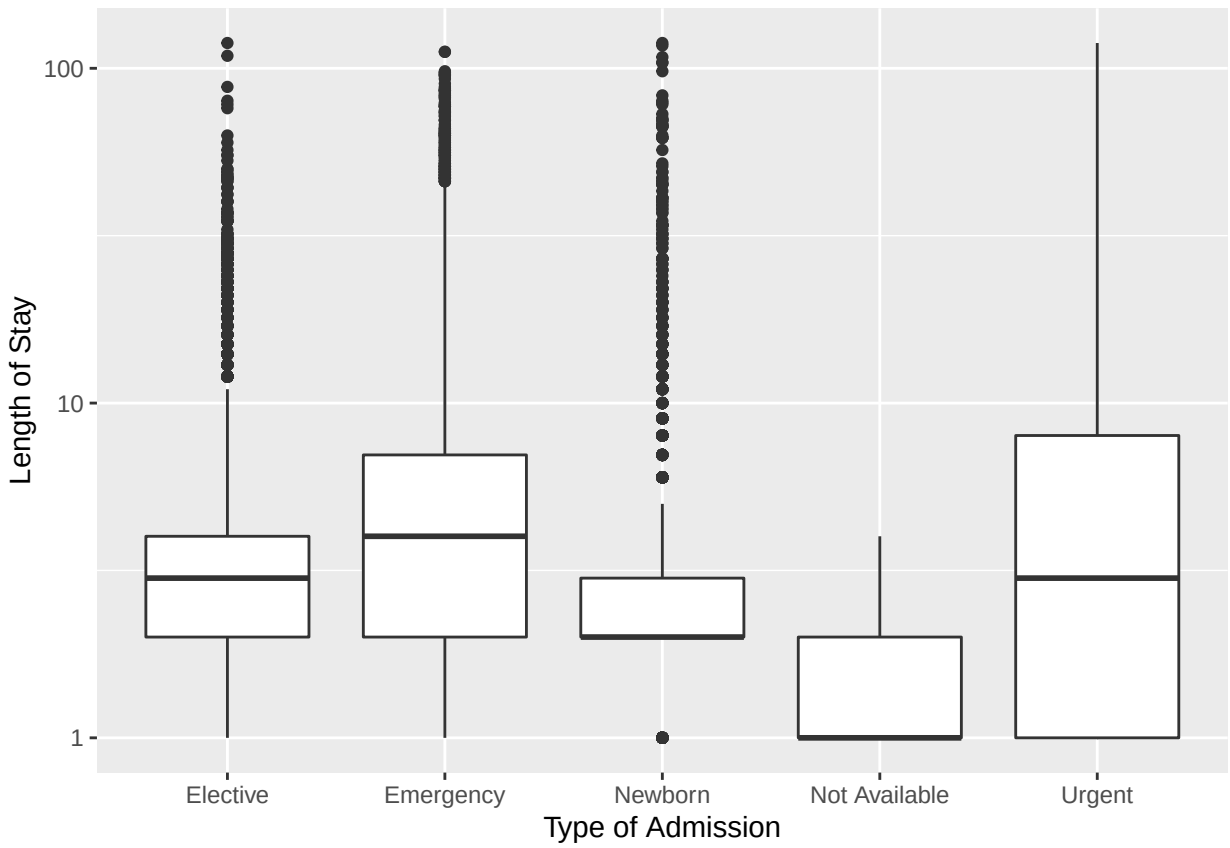

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Gender`, `Length of Stay`))
tmp.p1 <-tmp.p1 + geom_boxplot()+ scale_y_continuous(trans = "log10")
  # facet_grid(vars(`Type of Admission`))
 (tmp.p1)
```



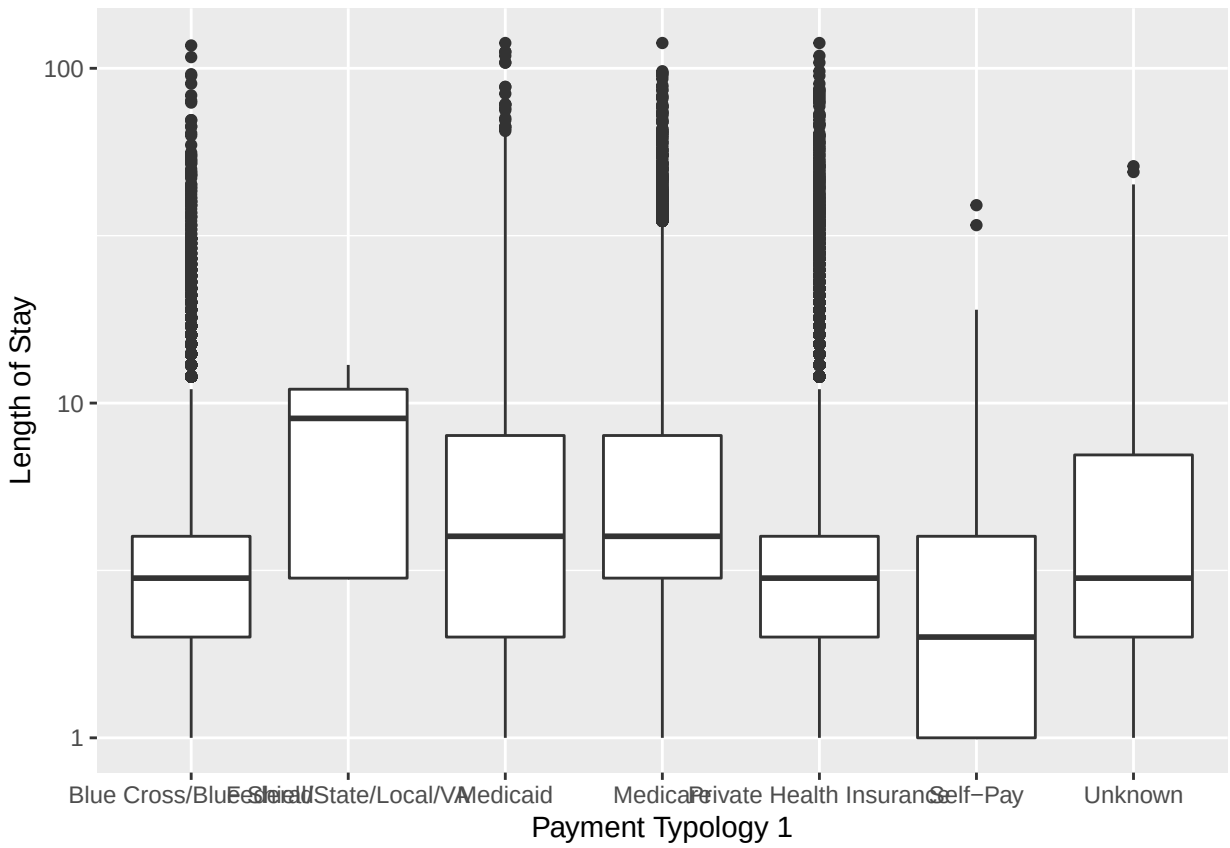

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Age Group`, `Length of Stay`))
tmp.p1 <-tmp.p1 + geom_boxplot()+ scale_y_continuous(trans = "log10")
  # facet_grid(vars(`Type of Admission`))
 (tmp.p1)
```



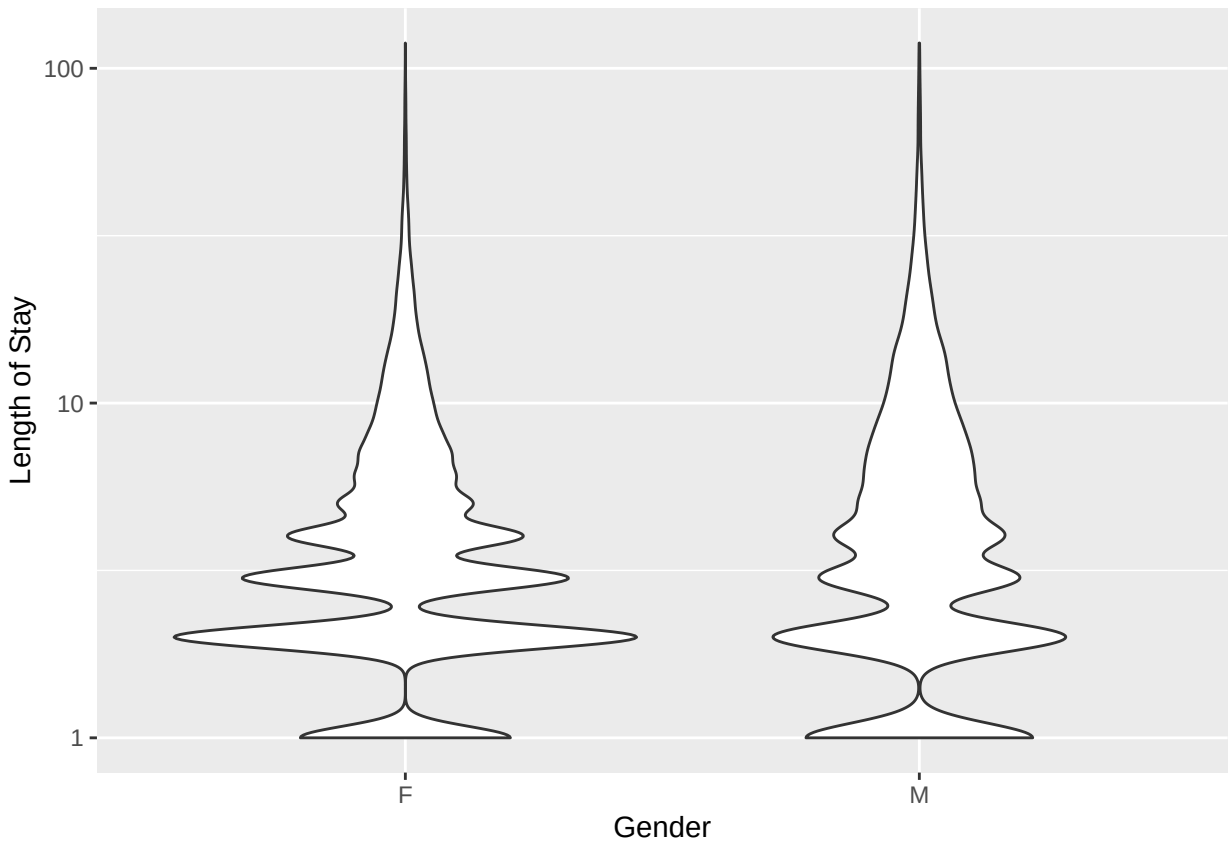

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Type of Admission`, `Length of Stay`))
tmp.p1 <-tmp.p1 + geom_boxplot()+ scale_y_continuous(trans = "log10")
  # facet_grid(vars(`Type of Admission`))
 (tmp.p1)
```



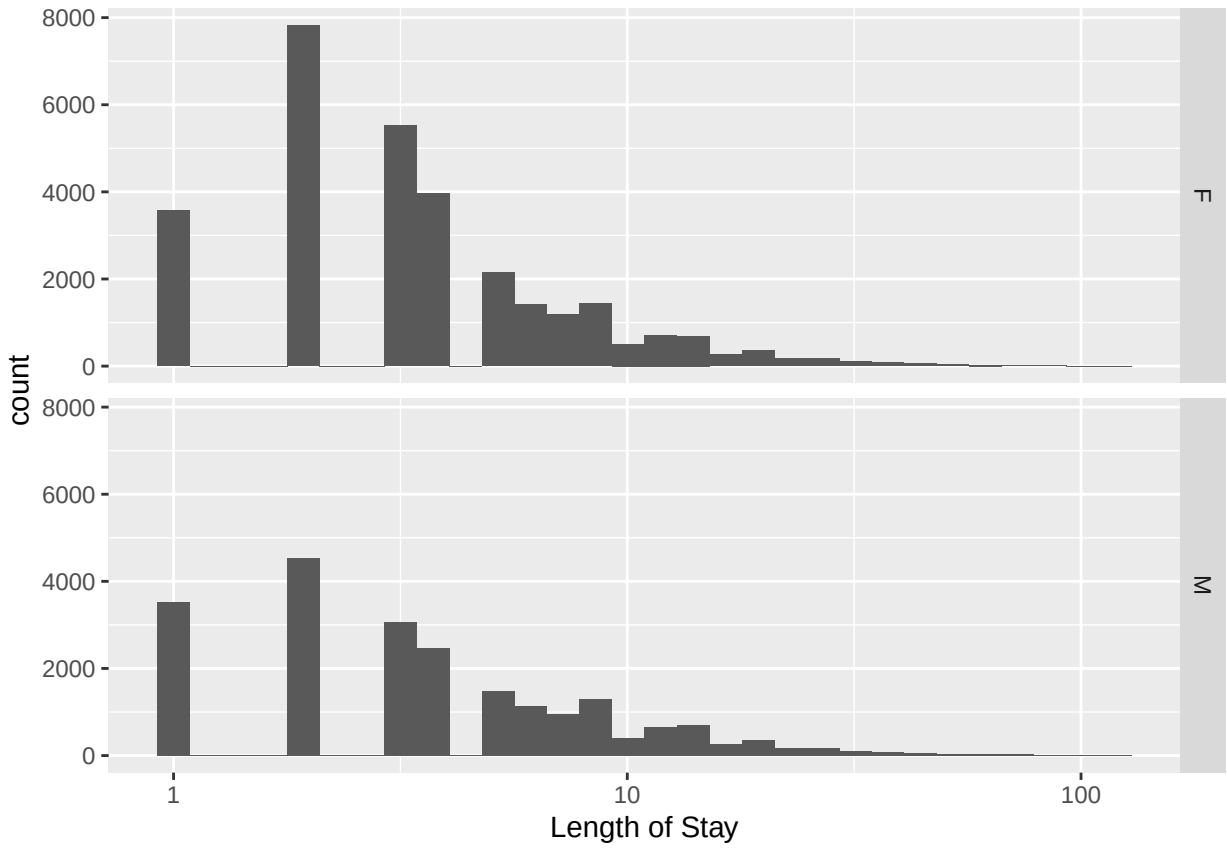

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Payment Typology 1`, `Length of Stay`))
tmp.p1 <-tmp.p1 + geom_boxplot()+ scale_y_continuous(trans = "log10")
  # facet_grid(vars(`Type of Admission`))
 (tmp.p1)
```



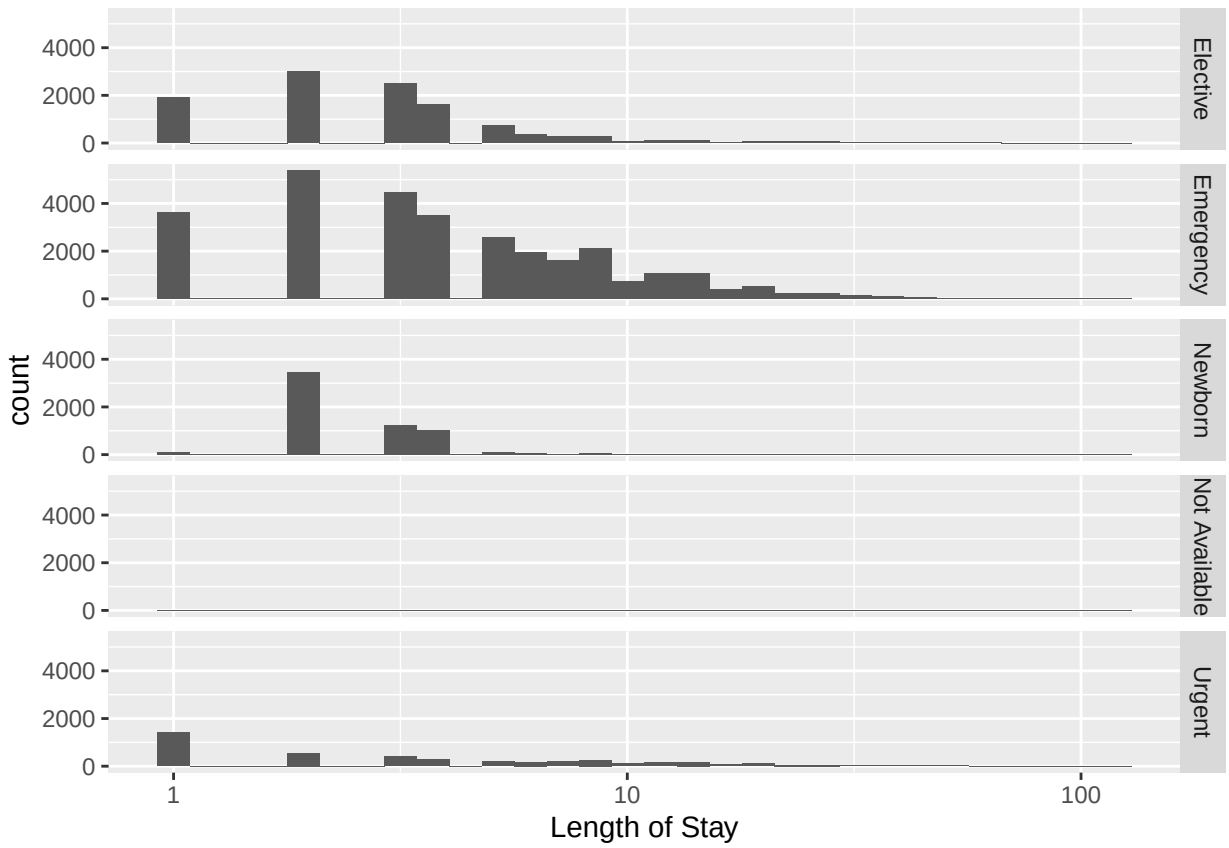

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Gender`, `Length of Stay`))
tmp.p1 <-tmp.p1 + geom_violin()+ scale_y_continuous(trans = "log10")
 (tmp.p1)
```



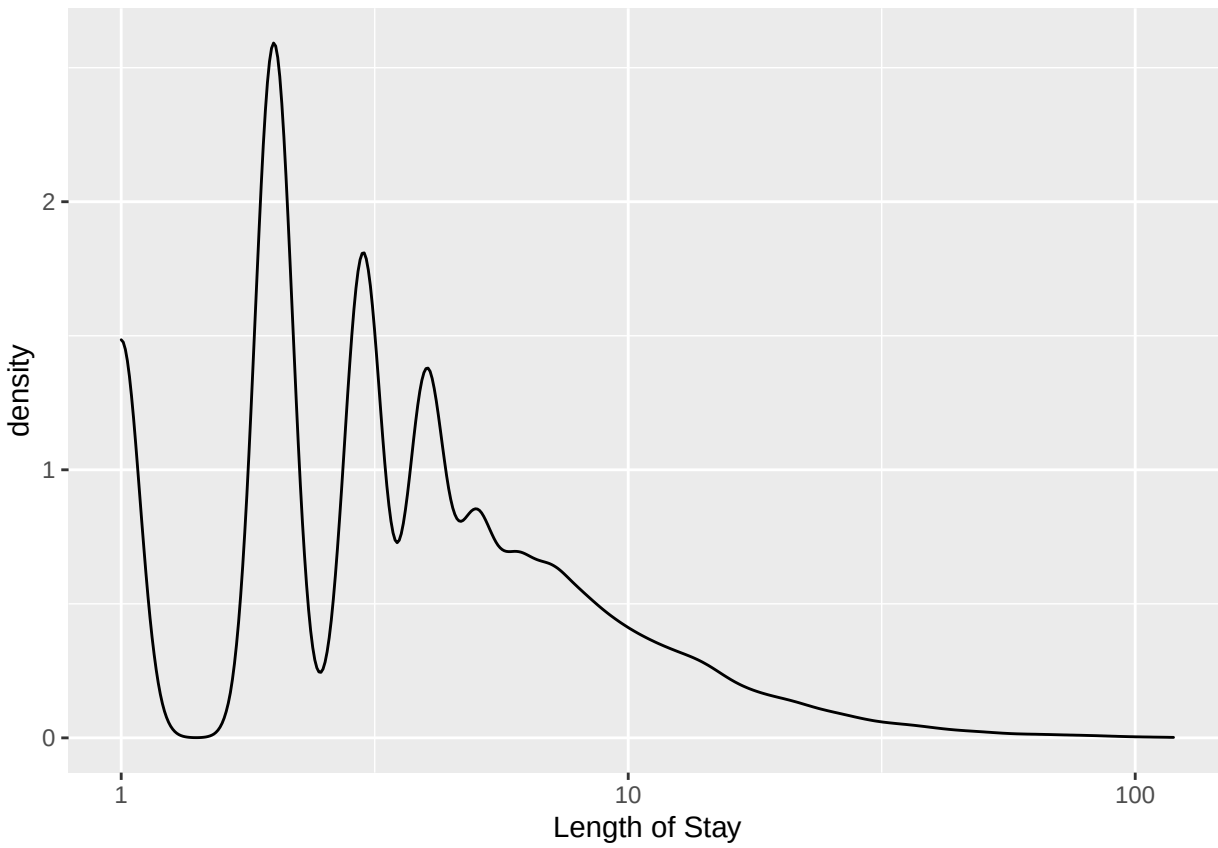

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Length of Stay`))
tmp.p1 <-tmp.p1 + geom_histogram()+facet_grid(vars(`Gender`))+
  scale_x_continuous(trans = "log10")
 (tmp.p1)
```




```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Length of Stay`))
tmp.p1 <-tmp.p1 + geom_histogram()+facet_grid(vars(`Type of Admission`))+
  scale_x_continuous(trans = "log10")
 (tmp.p1)
```




```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Length of Stay`))
tmp.p1 <- tmp.p1 + geom_density()+
  scale_x_continuous(trans = "log10")
 (tmp.p1)
```



Modelling Key Performance Indicators (KPI)

Regression models

In linear regression we have the following model:

$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ip}\beta_p = \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i$, i=1...N

- x: predicting (indepent) variables
- y: response (dependent) variable
- β: coefficients
- p: number of x-variables
- N: number of observations
- ϵ_i: residual ($\epsilon \sim N(0, \sigma_{error})$).

Predicting "Length of Stay" with linear regression model

First use the following predictors: * `Age Group` * `Gender` * `Type of Admission`

```
library(jtools)

tmp.m1<-glm(`Length of Stay`~`Age Group`+`Gender`+`Type of Admission`,
            data=eHeaTZ.NYSDOH.1)

summary(tmp.m1)
```

```
##
## Call:
## glm(formula = `Length of Stay` ~ `Age Group` + Gender + `Type of Admission`,
##     data = eHeaTZ.NYSDOH.1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
##  -6.405   -3.093   -1.368    0.632  115.241
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       1.82731    0.20979   8.710  < 2e-16 ***
## `Age Group`18 to 29               1.26646    0.22632   5.596 2.21e-08 ***
## `Age Group`30 to 49               1.37286    0.20892   6.571 5.04e-11 ***
## `Age Group`50 to 69               2.40161    0.20512  11.708  < 2e-16 ***
## `Age Group`70 or Older            2.87455    0.20365  14.115  < 2e-16 ***
## GenderM                           0.39134    0.06142   6.372 1.88e-10 ***
## `Type of Admission`Emergency      1.63377    0.07761  21.051  < 2e-16 ***
## `Type of Admission`Newborn        1.54027    0.22459   6.858 7.05e-12 ***
## `Type of Admission`Not Available -1.71647    3.83881  -0.447    0.655
## `Type of Admission`Urgent         2.31167    0.12312  18.776  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 44.18169)
##
##     Null deviance: 2379097  on 51850  degrees of freedom
## Residual deviance: 2290423  on 51841  degrees of freedom
##   (17 observations deleted due to missingness)
## AIC: 343586
##
## Number of Fisher Scoring iterations: 2
```

```
knitr::kable(ci.lin(tmp.m1),digits = 3)
```

	Estimate	StdErr	z	P	2.5%	97.5%
(Intercept)	1.827	0.210	8.710	0.000	1.416	2.238
`Age Group`18 to 29	1.266	0.226	5.596	0.000	0.823	1.710
`Age Group`30 to 49	1.373	0.209	6.571	0.000	0.963	1.782
`Age Group`50 to 69	2.402	0.205	11.708	0.000	2.000	2.804
`Age Group`70 or Older	2.875	0.204	14.115	0.000	2.475	3.274
GenderM	0.391	0.061	6.372	0.000	0.271	0.512
`Type of Admission`Emergency	1.634	0.078	21.051	0.000	1.482	1.786
`Type of Admission`Newborn	1.540	0.225	6.858	0.000	1.100	1.980
`Type of Admission`Not Available	-1.716	3.839	-0.447	0.655	-9.240	5.807
`Type of Admission`Urgent	2.312	0.123	18.776	0.000	2.070	2.553

Next, add `Payment Typology 1` to model as predictor.

```
tmp.m2<-update(tmp.m1,~.+`Payment Typology 1`)
summary(tmp.m2)
```

```
##
## Call:
```

```
## glm(formula = `Length of Stay` ~ `Age Group` + Gender + `Type of Admission` +
##     `Payment Typology 1`, data = eHeaTZ.NYSDOH.1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
##  -8.424   -3.023   -1.336    0.593  115.214
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                               1.26833    0.21880   5.797
## `Age Group`18 to 29                       1.27891    0.22514   5.681
## `Age Group`30 to 49                       1.45419    0.20809   6.988
## `Age Group`50 to 69                       2.14496    0.20682  10.371
## `Age Group`70 or Older                    1.75760    0.22281   7.888
## GenderM                                   0.34229    0.06115   5.597
## `Type of Admission`Emergency              1.50879    0.07807  19.325
## `Type of Admission`Newborn                1.72562    0.22382   7.710
## `Type of Admission`Not Available         -2.08136    3.81547  -0.546
## `Type of Admission`Urgent                 2.14582    0.12281  17.473
## `Payment Typology 1`Federal/State/Local/VA    2.36002    2.95599   0.798
## `Payment Typology 1`Medicaid              3.52239    0.16484  21.368
## `Payment Typology 1`Medicare              1.87244    0.11227  16.678
## `Payment Typology 1`Private Health Insurance  0.44934    0.08247   5.449
## `Payment Typology 1`Self-Pay             -0.66454    0.48412  -1.373
## `Payment Typology 1`Unknown               1.06642    0.27736   3.845
##                                          Pr(>|t|)
## (Intercept)                              6.79e-09 ***
## `Age Group`18 to 29                      1.35e-08 ***
## `Age Group`30 to 49                      2.82e-12 ***
## `Age Group`50 to 69                       < 2e-16 ***
## `Age Group`70 or Older                   3.12e-15 ***
## GenderM                                  2.19e-08 ***
## `Type of Admission`Emergency              < 2e-16 ***
## `Type of Admission`Newborn               1.28e-14 ***
## `Type of Admission`Not Available         0.585408
## `Type of Admission`Urgent                 < 2e-16 ***
## `Payment Typology 1`Federal/State/Local/VA   0.424650
## `Payment Typology 1`Medicaid              < 2e-16 ***
## `Payment Typology 1`Medicare              < 2e-16 ***
## `Payment Typology 1`Private Health Insurance 5.10e-08 ***
## `Payment Typology 1`Self-Pay             0.169864
## `Payment Typology 1`Unknown              0.000121 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 43.64376)
##
##     Null deviance: 2379097  on 51850  degrees of freedom
## Residual deviance: 2262274  on 51835  degrees of freedom
##   (17 observations deleted due to missingness)
## AIC: 342957
##
## Number of Fisher Scoring iterations: 2
```

Check, if `Payment Typology 1` improved the model.

```
anova(tmp.m1,tmp.m2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: `Length of Stay` ~ `Age Group` + Gender + `Type of Admission`
## Model 2: `Length of Stay` ~ `Age Group` + Gender + `Type of Admission` +
##     `Payment Typology 1`
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1     51841    2290423
## 2     51835    2262274  6    28149 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
knitr::kable(ci.lin(tmp.m2),digits = 3)
```

	Estimate	StdErr	z	P	2.5%	97.5%
(Intercept)	1.268	0.219	5.797	0.000	0.840	1.697
`Age Group`18 to 29	1.279	0.225	5.681	0.000	0.838	1.720
`Age Group`30 to 49	1.454	0.208	6.988	0.000	1.046	1.862
`Age Group`50 to 69	2.145	0.207	10.371	0.000	1.740	2.550
`Age Group`70 or Older	1.758	0.223	7.888	0.000	1.321	2.194
GenderM	0.342	0.061	5.597	0.000	0.222	0.462
`Type of Admission`Emergency	1.509	0.078	19.325	0.000	1.356	1.662
`Type of Admission`Newborn	1.726	0.224	7.710	0.000	1.287	2.164
`Type of Admission`Not Available	-2.081	3.815	-0.546	0.585	-9.560	5.397
`Type of Admission`Urgent	2.146	0.123	17.473	0.000	1.905	2.387
`Payment Typology 1`Federal/State/Local/VA	2.360	2.956	0.798	0.425	-3.434	8.154
`Payment Typology 1`Medicaid	3.522	0.165	21.368	0.000	3.199	3.845
`Payment Typology 1`Medicare	1.872	0.112	16.678	0.000	1.652	2.092
`Payment Typology 1`Private Health Insurance	0.449	0.082	5.449	0.000	0.288	0.611
`Payment Typology 1`Self-Pay	-0.665	0.484	-1.373	0.170	-1.613	0.284
`Payment Typology 1`Unknown	1.066	0.277	3.845	0.000	0.523	1.610