Example of data analyses

Jari Haukka

2021-08-31

Contents

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

```
# 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
```

```r
# 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`=="Mount Sinai Hospital")#2
# save(eHeaTZ.NYSDOH.1,file="eHeaTZ.NYSDOH.1.RData")

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

Checking data with simple summaries

We use "Mount Sinai 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
## 1651002         New York City       Manhattan                     7002024
## 1651003         New York City       Manhattan                     7002024
## 1651004         New York City       Manhattan                     7002024
## 1651005         New York City       Manhattan                     7002024
## 1651006         New York City       Manhattan                     7002024
##         Permanent Facility Id        Facility Name Age Group
## 1651002                001456 Mount Sinai Hospital  50 to 69
## 1651003                001456 Mount Sinai Hospital  18 to 29
## 1651004                001456 Mount Sinai Hospital  18 to 29
## 1651005                001456 Mount Sinai Hospital  18 to 29
## 1651006                001456 Mount Sinai Hospital  18 to 29
##         Zip Code - 3 digits Gender                   Race         Ethnicity
## 1651002                 105      M                  White Not Span/Hispanic
## 1651003                 100      F Black/African American Not Span/Hispanic
## 1651004                 100      F             Other Race Not Span/Hispanic
## 1651005                 100      F             Other Race Not Span/Hispanic
## 1651006                 112      F Black/African American Not Span/Hispanic
##         Length of Stay Type of Admission Patient Disposition Discharge Year
## 1651002              1          Elective   Home or Self Care           2012
## 1651003              3          Elective   Home or Self Care           2012
## 1651004              3          Elective   Home or Self Care           2012
## 1651005              4          Elective   Home or Self Care           2012
## 1651006              2          Elective   Home or Self Care           2012
##         CCS Diagnosis Code
## 1651002                106
## 1651003                189
## 1651004                191
## 1651005                195
## 1651006                191
##                                                      CCS Diagnosis Description
## 1651002                                                   Cardiac dysrhythmias
## 1651003                                                     Previous C-section
## 1651004                     Polyhydramnios and other problems of amniotic cavity
```

```
## 1651005 Other complications of birth; puerperium affecting management of mother
## 1651006                    Polyhydramnios and other problems of amniotic cavity
##         CCS Procedure Code CCS Procedure Description APR DRG Code
## 1651002                049       OTHER OR HEART PRCS          175
## 1651003                134          CESAREAN SECTION          540
## 1651004                137   OT PRCS TO ASSIST DELIV          560
## 1651005                134          CESAREAN SECTION          540
## 1651006                137   OT PRCS TO ASSIST DELIV          560
##                                APR DRG Description APR MDC Code
## 1651002 Percutaneous coronary intervention w/o AMI           05
## 1651003                          Cesarean delivery           14
## 1651004                           Vaginal delivery           14
## 1651005                          Cesarean delivery           14
## 1651006                           Vaginal delivery           14
##                                    APR MDC Description
## 1651002 Diseases and Disorders of the Circulatory System
## 1651003         Pregnancy, Childbirth and the Puerperium
## 1651004         Pregnancy, Childbirth and the Puerperium
## 1651005         Pregnancy, Childbirth and the Puerperium
## 1651006         Pregnancy, Childbirth and the Puerperium
##         APR Severity of Illness Code APR Severity of Illness Description
## 1651002                            1                                Minor
## 1651003                            1                                Minor
## 1651004                            3                                Major
## 1651005                            1                                Minor
## 1651006                            1                                Minor
##         APR Risk of Mortality APR Medical Surgical Description
## 1651002                 Minor                         Surgical
## 1651003                 Minor                         Surgical
## 1651004                 Minor                          Medical
## 1651005                 Minor                         Surgical
## 1651006                 Minor                          Medical
##               Payment Typology 1 Payment Typology 2 Payment Typology 3
## 1651002 Private Health Insurance           Self-Pay               <NA>
## 1651003                 Medicaid           Medicaid           Self-Pay
## 1651004                 Medicaid           Medicaid           Self-Pay
## 1651005                 Medicaid           Self-Pay               <NA>
## 1651006                 Medicaid           Medicaid           Self-Pay
##         Birth Weight Abortion Edit Indicator Emergency Department Indicator
## 1651002         0000                       N                              N
## 1651003         0000                       N                              N
## 1651004         0000                       N                              N
## 1651005         0000                       N                              N
## 1651006         0000                       N                              N
##         Total Charges Total Costs Ratio of Total Costs to Total Charges
## 1651002      52756.15    15896.84                            0.3013268
## 1651003      22328.95    11074.50                            0.4959704
## 1651004      21583.00    10572.70                            0.4898626
## 1651005      27540.80    13661.81                            0.4960572
## 1651006      15129.28     7428.46                            0.4909986
```

Descriptive tables / `onetable`-package

```
# 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)
```

```
# 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	32438	26284			
Age Group (%)			<0.001		0.0
0 to 17	4701 (14.5)	5294 (20.1)			
18 to 29	3950 (12.2)	1422 (5.4)			
30 to 49	8782 (27.1)	3915 (14.9)			
50 to 69	7636 (23.5)	9454 (36.0)			
70 or Older	7369 (22.7)	6199 (23.6)			
Gender = M (%)	0 (0.0)	26284 (100.0)	<0.001		0.0
Payment Typology 1 (%)			<0.001		0.0
Blue Cross/Blue Shield	3445 (10.6)	2922 (11.1)			
Federal/State/Local/VA	23 (0.1)	8 (0.0)			
Managed Care, Unspecified	181 (0.6)	179 (0.7)			
Medicaid	8069 (24.9)	6574 (25.0)			
Medicare	10497 (32.4)	9778 (37.2)			
Miscellaneous/Other	152 (0.5)	275 (1.0)			
Private Health Insurance	9714 (29.9)	6047 (23.0)			

	F	M	p	test	Missing
Self-Pay	357 (1.1)	501 (1.9)			
Emergency Department Indicator = Y (%)	10885 (33.6)	10130 (38.5)	<0.001		0.0
Length of Stay (mean (SD))	5.19 (7.64)	5.91 (9.03)	<0.001		0.1
Type of Admission (%)			<0.001		0.0
Elective	14776 (45.6)	8657 (32.9)			
Emergency	11002 (33.9)	10267 (39.1)			
Newborn	3230 (10.0)	3499 (13.3)			
Not Available	201 (0.6)	165 (0.6)			
Urgent	3229 (10.0)	3696 (14.1)			
APR Severity of Illness Description (%)			<0.001		0.0
Extreme	1935 (6.0)	2402 (9.1)			
Major	6658 (20.5)	6601 (25.1)			
Minor	11816 (36.4)	8029 (30.6)			
Moderate	12029 (37.1)	9249 (35.2)			
Total Charges (mean (SD))	41221.65 (68508.22)	51557.32 (88634.46)	<0.001		0.0
Total Costs (mean (SD))	17026.08 (29549.11)	20604.18 (36635.97)	<0.001		0.0
Ratio of Total Costs to Total Charges (mean (SD))	0.41 (0.11)	0.39 (0.11)	<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	23433	21269	6729	366	6925			
Age Group (%)						<0.001		0.0
0 to 17	901 (3.8)	1869 (8.8)	6709 (99.7)	37 (10.1)	479 (6.9)			
18 to 29	3149 (13.4)	1843 (8.7)	11 (0.2)	27 (7.4)	342 (4.9)			
30 to 49	7483 (31.9)	4038 (19.0)	9 (0.1)	63 (17.2)	1104 (15.9)			
50 to 69	7221 (30.8)	7133 (33.5)	0 (0.0)	140 (38.3)	2596 (37.5)			
70 or Older	4679 (20.0)	6386 (30.0)	0 (0.0)	99 (27.0)	2404 (34.7)			
Gender = M (%)	8657 (36.9)	10267 (48.3)	3499 (52.0)	165 (45.1)	3696 (53.4)	<0.001		0.0
Payment Typology 1 (%)						<0.001		0.0
Blue Cross/Blue Shield	3374 (14.4)	1399 (6.6)	850 (12.6)	47 (12.8)	697 (10.1)			
Federal/State/Local/VA	13 (0.1)	9 (0.0)	1 (0.0)	0 (0.0)	8 (0.1)			
Managed Care, Unspecified	80 (0.3)	218 (1.0)	30 (0.4)	1 (0.3)	31 (0.4)			
Medicaid	4487 (19.1)	6580 (30.9)	2060 (30.6)	69 (18.9)	1447 (20.9)			
Medicare	7013 (29.9)	9679 (45.5)	0 (0.0)	148 (40.4)	3435 (49.6)			

	Elective	Emergency	Newborn	Not Available	Urgent	p	test	Missing
Miscellaneous/Other	290 (1.2)	64 (0.3)	0 (0.0)	0 (0.0)	73 (1.1)			
Private Health Insurance	7890 (33.7)	2889 (13.6)	3727 (55.4)	99 (27.0)	1156 (16.7)			
Self-Pay	286 (1.2)	431 (2.0)	61 (0.9)	2 (0.5)	78 (1.1)			
Emergency Department Indicator = Y (%)	96 (0.4)	20865 (98.1)	4 (0.1)	5 (1.4)	45 (0.6)	<0.001		0.0
Length of Stay (mean (SD))	3.83 (5.54)	6.40 (8.74)	3.85 (8.02)	8.61 (11.69)	9.95 (11.89)	<0.001		0.1
Type of Admission (%)						<0.001		0.0
Elective	23433 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)			
Emergency	0 (0.0)	21269 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)			
Newborn	0 (0.0)	0 (0.0)	6729 (100.0)	0 (0.0)	0 (0.0)			
Not Available	0 (0.0)	0 (0.0)	0 (0.0)	366 (100.0)	0 (0.0)			
Urgent	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	6925 (100.0)			
APR Severity of Illness Description (%)						<0.001		0.0
Extreme	661 (2.8)	2309 (10.9)	73 (1.1)	52 (14.2)	1242 (17.9)			
Major	3629 (15.5)	6809 (32.0)	298 (4.4)	118 (32.2)	2405 (34.7)			
Minor	9807 (41.9)	3993 (18.8)	5125 (76.2)	67 (18.3)	853 (12.3)			
Moderate	9336 (39.8)	8158 (38.4)	1230 (18.3)	129 (35.2)	2425 (35.0)			
Total Charges (mean (SD))	40224.55 (56505.81)	47079.14 (72771.95)	19783.24 (86605.63)	90125.26 (135200.32)	84081.60 (119878.60)	<0.001		0.0
Total Costs (mean (SD))	16896.12 (22504.60)	19125.98 (30230.83)	9006.83 (44804.83)	32733.56 (43287.48)	31559.22 (48610.81)	<0.001		0.0
Ratio of Total Costs to Total Charges (mean (SD))	0.44 (0.08)	0.41 (0.10)	0.25 (0.09)	0.40 (0.09)	0.38 (0.10)	<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	Managed Care, Unspecified	Medicaid	Medicare	Miscellaneous/Other	Private Health Insurance	Self-Pay	p	test	Missing
n	6367	31	360	14643	20275	427	15761	858			
Age Group (%)									<0.001		0.0
0 to 17	1180 (18.5)	7 (22.6)	63 (17.5)	4070 (27.8)	11 (0.1)	2 (0.5)	4571 (29.0)	91 (10.6)			

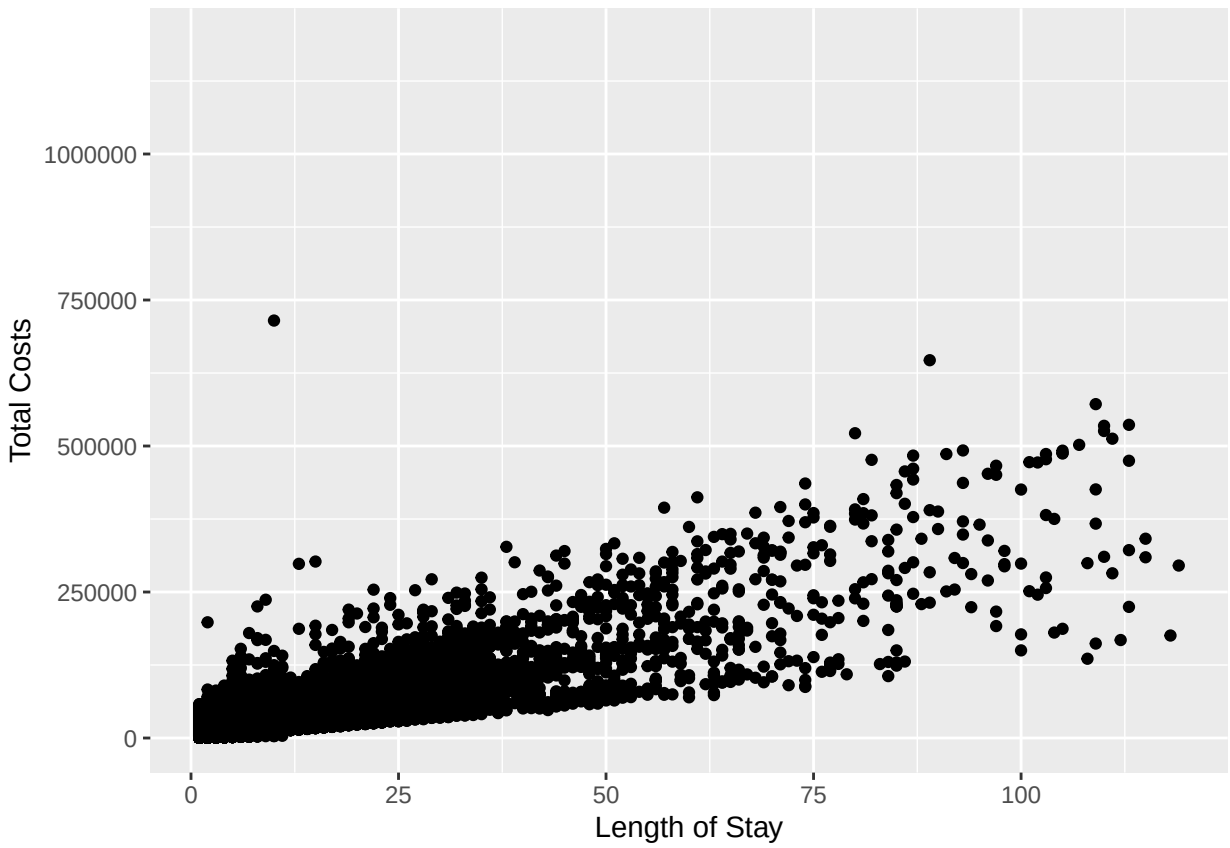
	Blue Cross/Blue Shield	Federal/State/Local/VA	Managed Care, Unspecified	Medicaid	Medicare	Miscellaneous/Other	Private Health Insurance	Self-Pay	p	test	Missing
18 to 29	587 (9.2)	8 (25.8)	51 (14.2)	2666 (18.2)	248 (1.2)	45 (10.5)	1613 (10.2)	154 (17.9)			
30 to 49	2008 (31.5)	6 (19.4)	99 (27.5)	3594 (24.5)	1160 (5.7)	162 (37.9)	5388 (34.2)	280 (32.6)			
50 to 69	2427 (38.1)	10 (32.3)	137 (38.1)	4008 (27.4)	6198 (30.6)	200 (46.8)	3819 (24.2)	291 (33.9)			
70 or Older	165 (2.6)	0 (0.0)	10 (2.8)	305 (2.1)	12658 (62.4)	18 (4.2)	370 (2.3)	42 (4.9)			
Gender = M (%)	2922 (45.9)	8 (25.8)	179 (49.7)	6574 (44.9)	9778 (48.2)	275 (64.4)	6047 (38.4)	501 (58.4)	<0.001		0.0
Payment Typology 1 (%)									<0.001		0.0
Blue Cross/Blue Shield	6367 (100.0)	0 (0.0)	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)	31 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)			
Managed Care, Unspecified	0 (0.0)	0 (0.0)	360 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)			
Medicaid	0 (0.0)	0 (0.0)	0 (0.0)	14643 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)			
Medicare	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	20275 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)			
Miscellaneous/Other	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	427 (100.0)	0 (0.0)	0 (0.0)			
Private Health Insurance	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	15761 (100.0)	0 (0.0)			
Self-Pay	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	858 (100.0)			
Emergency Department Indicator = Y (%)	1385 (21.8)	9 (29.0)	215 (59.7)	6437 (44.0)	9608 (47.4)	65 (15.2)	2862 (18.2)	434 (50.6)	<0.001		0.0
Length of Stay (mean (SD))	4.81 (7.59)	4.74 (6.63)	5.61 (10.31)	5.96 (9.14)	6.53 (8.92)	5.19 (6.79)	4.17 (6.60)	3.94 (6.89)	<0.001		0.1
Type of Admission (%)									<0.001		0.0
Elective	3374 (53.0)	13 (41.9)	80 (22.2)	4487 (30.6)	7013 (34.6)	290 (67.9)	7890 (50.1)	286 (33.3)			
Emergency	1399 (22.0)	9 (29.0)	218 (60.6)	6580 (44.9)	9679 (47.7)	64 (15.0)	2889 (18.3)	431 (50.2)			
Newborn	850 (13.4)	1 (3.2)	30 (8.3)	2060 (14.1)	0 (0.0)	0 (0.0)	3727 (23.6)	61 (7.1)			
Not Available	47 (0.7)	0 (0.0)	1 (0.3)	69 (0.5)	148 (0.7)	0 (0.0)	99 (0.6)	2 (0.2)			
Urgent	697 (10.9)	8 (25.8)	31 (8.6)	1447 (9.9)	3435 (16.9)	73 (17.1)	1156 (7.3)	78 (9.1)			

	Blue Cross/Blue Shield	Federal/State/Local/VA	Managed Care, Unspecified	Medicaid	Medicare	Miscellaneous/Other	Private/Commercial Health Insurance	Self-Pay	p	test	Missing
APR Severity of Illness Description (%)									<0.001		0.0
Extreme	384 (6.0)	3 (9.7)	34 (9.4)	984 (6.7)	2331 (11.5)	15 (3.5)	552 (3.5)	34 (4.0)			
Major	1074 (16.9)	3 (9.7)	81 (22.5)	3093 (21.1)	6910 (34.1)	55 (12.9)	1946 (12.3)	97 (11.3)			
Minor	2735 (43.0)	8 (25.8)	117 (32.5)	5008 (34.2)	3269 (16.1)	212 (49.6)	8089 (51.3)	407 (47.5)			
Moderate	2174 (34.1)	17 (54.8)	128 (35.6)	5558 (38.0)	7765 (38.3)	145 (34.0)	5172 (32.8)	319 (37.2)			
Total Charges (mean (SD))	44642.95 (78461.04)	51949.83 (82041.27)	49240.17 (96746.08)	43909.87 (81556.27)	57165.68 (84215.02)	52331.81 (78070.23)	33908.97 (62889.29)	34860.84 (86277.40)	<0.001		0.0
Total Costs (mean (SD))	18586.87 (34061.11)	21918.60 (36884.07)	20052.22 (44002.58)	19064.18 (38012.65)	21599.13 (31567.06)	22690.01 (35160.95)	14542.15 (28421.82)	13572.17 (25820.89)	<0.001		0.0
Ratio of Total Costs to Total Charges (mean (SD))	0.41 (0.11)	0.43 (0.13)	0.40 (0.11)	0.42 (0.14)	0.38 (0.07)	0.44 (0.10)	0.40 (0.13)	0.40 (0.10)	<0.001		0.0

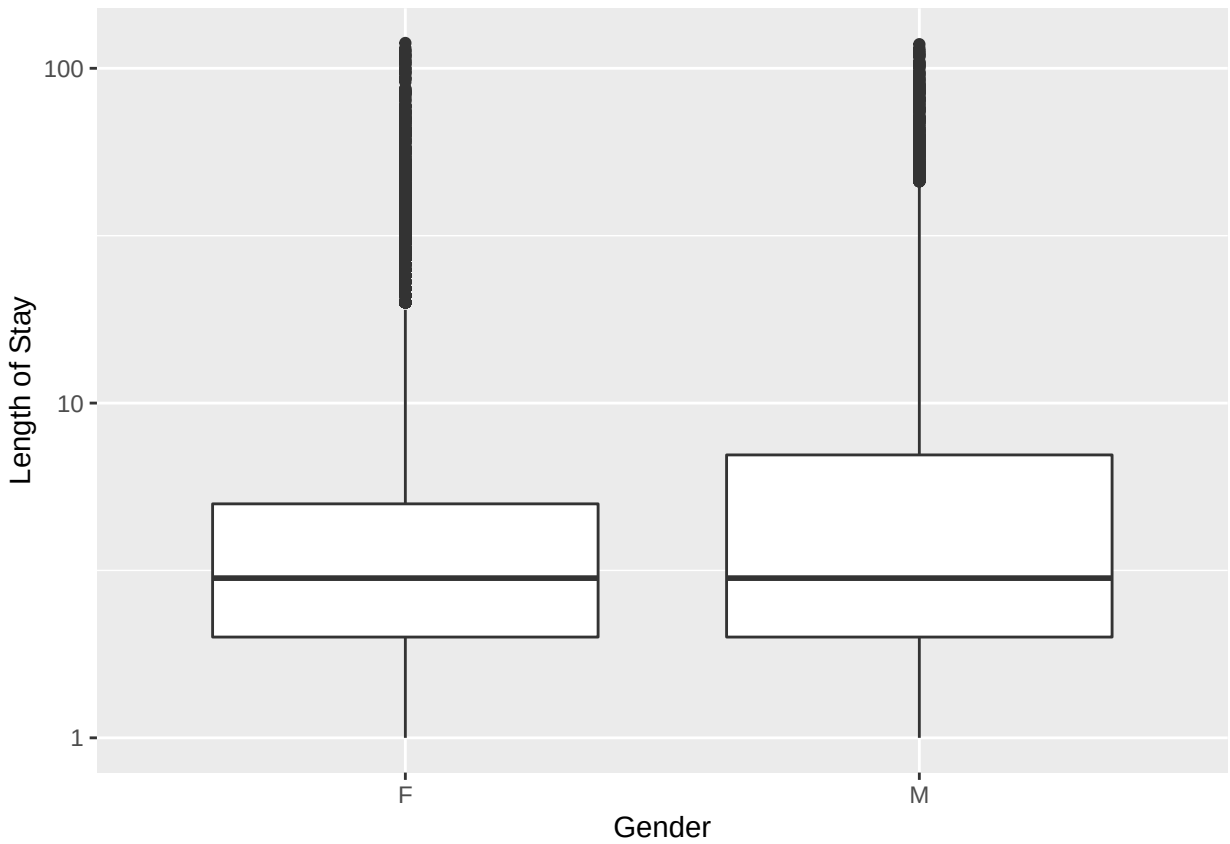
Visualization

Plot types

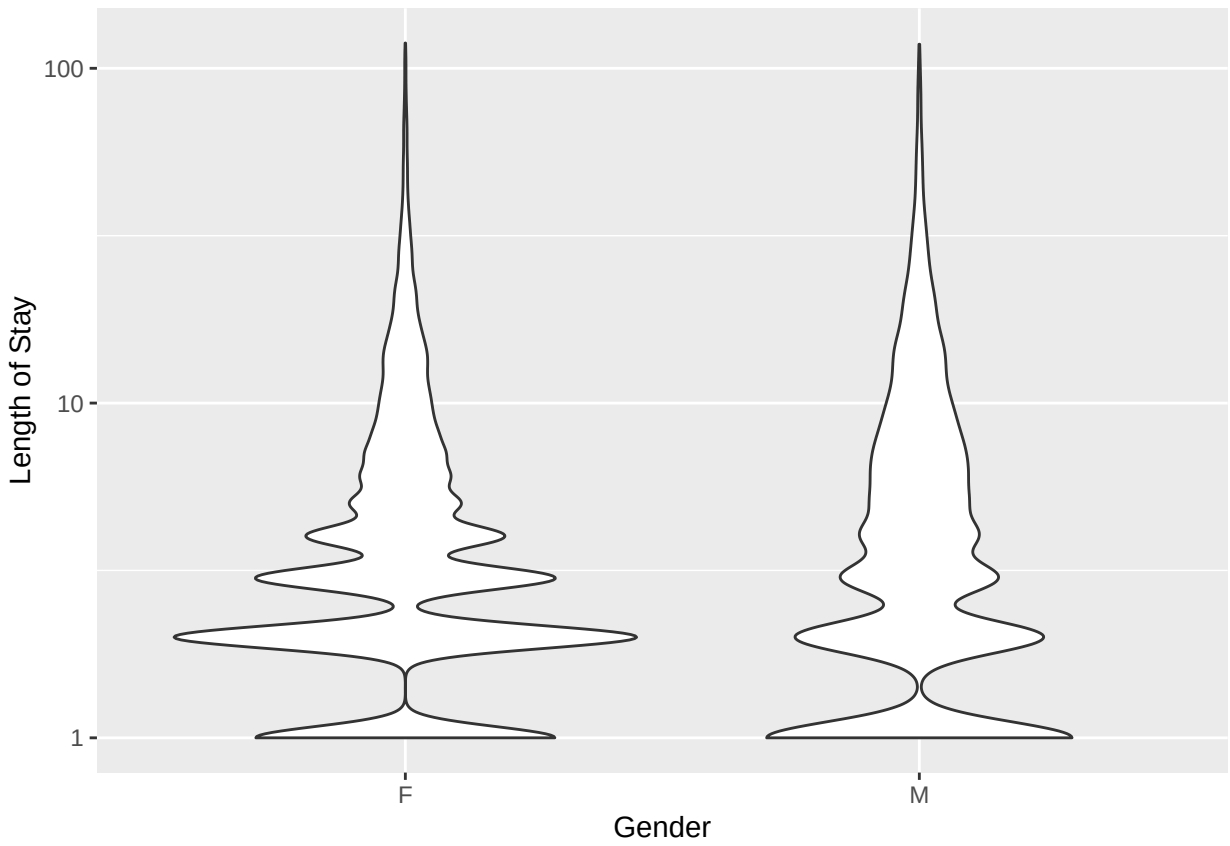
```
library(ggplot2)
library(plotly)
library(hrbrthemes)



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



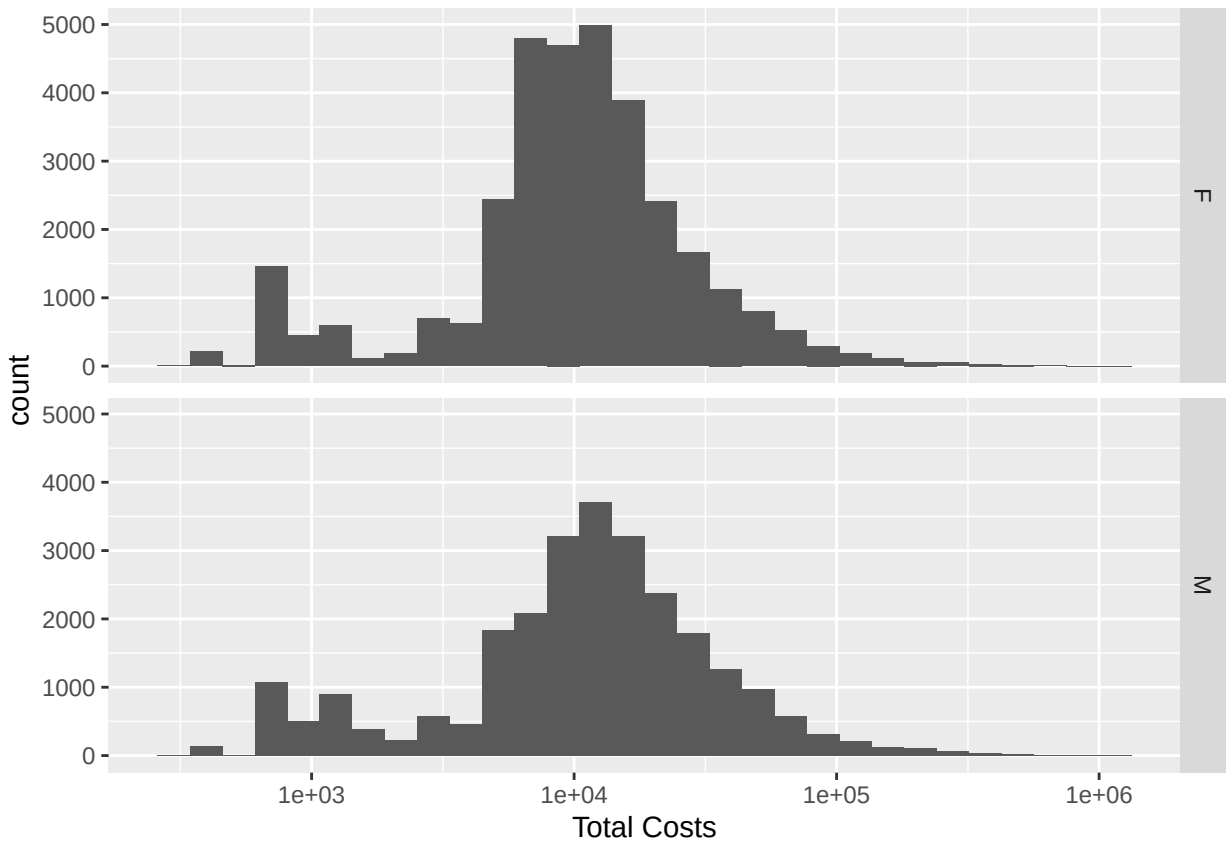

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Gender`, `Length of Stay`))
tmp.p1 <-tmp.p1 + geom_boxplot()+ scale_y_continuous(trans = "log10")
  # facet_grid(vars(`Type of Admission`))
 (tmp.p1)
```



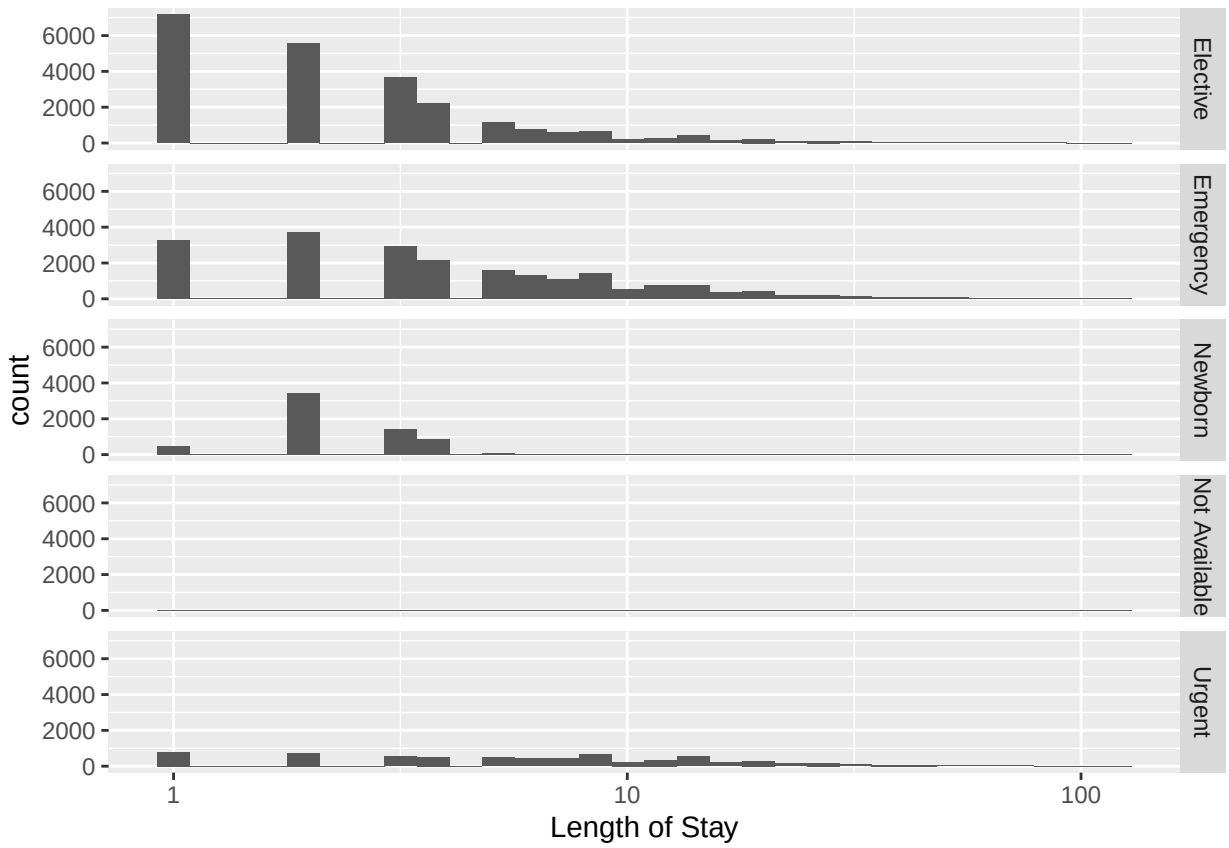

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Gender`, `Length of Stay`))
tmp.p1 <-tmp.p1 + geom_violin()+ scale_y_continuous(trans = "log10")
(tmp.p1)
```



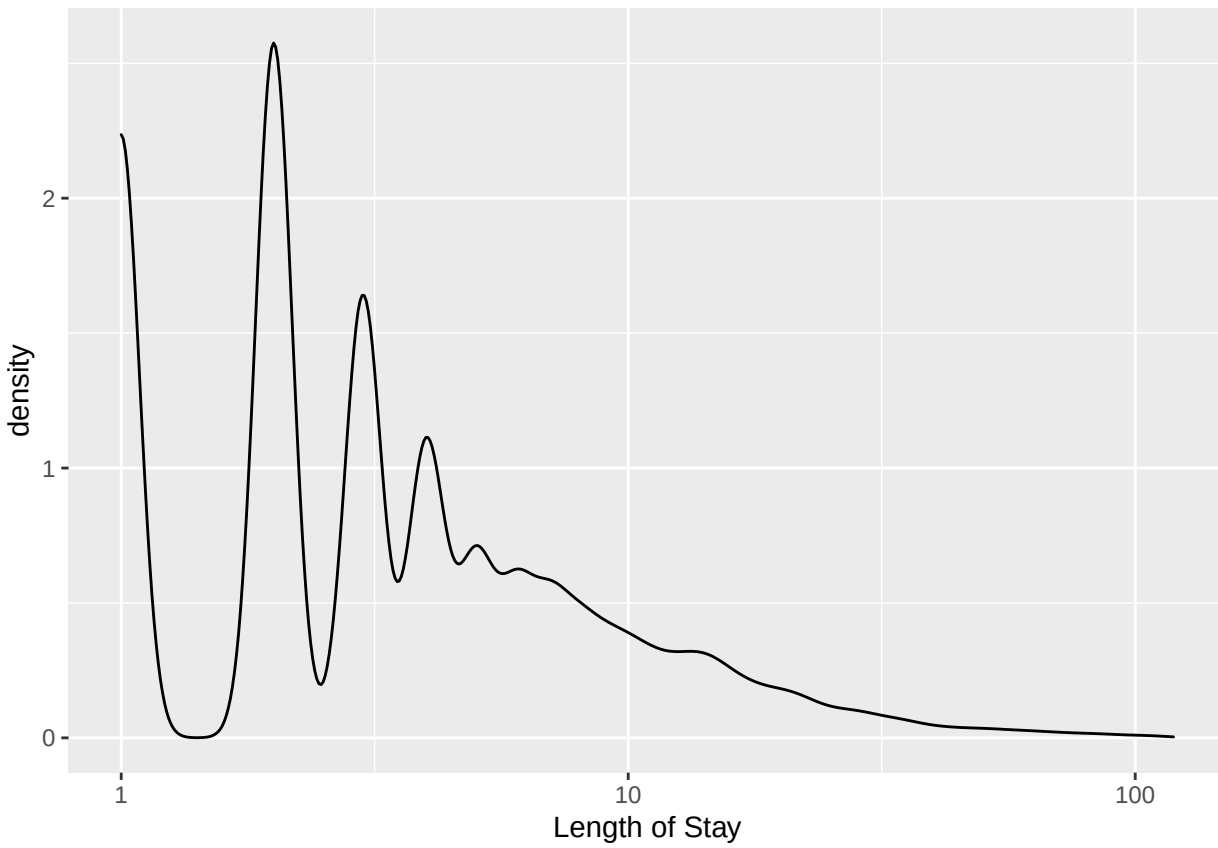

```
tmp.p1 <- ggplot(eHeaTZ.NYSDOH.1, aes(`Total Costs`))
tmp.p1 <-tmp.p1 + geom_histogram()+facet_grid(vars(`Gender`))+
  scale_x_continuous(trans = "log10")
(tmp.p1)
```




```
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
##  -9.311   -3.109   -1.684    0.389  112.504
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.30809    0.15325  21.586  < 2e-16 ***
## `Age Group`18 to 29               0.23603    0.17999   1.311  0.18976
## `Age Group`30 to 49               0.25770    0.15942   1.616  0.10600
## `Age Group`50 to 69               0.48306    0.15340   3.149  0.00164 **
## `Age Group`70 or Older            0.68627    0.15643   4.387 1.15e-05 ***
## GenderM                           0.31750    0.06871   4.621 3.82e-06 ***
## `Type of Admission`Emergency      2.50193    0.07786  32.133  < 2e-16 ***
## `Type of Admission`Newborn        0.37550    0.17859   2.103  0.03551 *
## `Type of Admission`Not Available  4.72786    0.42476  11.131  < 2e-16 ***
## `Type of Admission`Urgent         5.99921    0.11213  53.504  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 64.73779)
##
##     Null deviance: 4044001  on 58686  degrees of freedom
## Residual deviance: 3798620  on 58677  degrees of freedom
##   (35 observations deleted due to missingness)
## AIC: 411304
##
## Number of Fisher Scoring iterations: 2
```

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

	Estimate	StdErr	z	P	2.5%	97.5%
(Intercept)	3.308	0.153	21.586	0.000	3.008	3.608
`Age Group`18 to 29	0.236	0.180	1.311	0.190	-0.117	0.589
`Age Group`30 to 49	0.258	0.159	1.616	0.106	-0.055	0.570
`Age Group`50 to 69	0.483	0.153	3.149	0.002	0.182	0.784
`Age Group`70 or Older	0.686	0.156	4.387	0.000	0.380	0.993
GenderM	0.318	0.069	4.621	0.000	0.183	0.452
`Type of Admission`Emergency	2.502	0.078	32.133	0.000	2.349	2.655
`Type of Admission`Newborn	0.375	0.179	2.103	0.036	0.025	0.726
`Type of Admission`Not Available	4.728	0.425	11.131	0.000	3.895	5.560
`Type of Admission`Urgent	5.999	0.112	53.504	0.000	5.779	6.219

```
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")
```

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

```r
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
##  -9.585   -3.402   -1.632    0.456  112.526
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                                 3.09123    0.18177  17.006
## `Age Group`18 to 29                         0.26227    0.18001   1.457
## `Age Group`30 to 49                         0.43162    0.16107   2.680
## `Age Group`50 to 69                         0.40995    0.15917   2.575
## `Age Group`70 or Older                      0.19282    0.18481   1.043
## GenderM                                     0.30629    0.06877   4.454
## `Type of Admission`Emergency                2.29746    0.07984  28.776
## `Type of Admission`Newborn                  0.56543    0.17938   3.152
## `Type of Admission`Not Available            4.67529    0.42405  11.025
## `Type of Admission`Urgent                   5.86394    0.11245  52.148
## `Payment Typology 1`Federal/State/Local/VA -0.91033    1.44653  -0.629
## `Payment Typology 1`Managed Care, Unspecified  0.09407    0.43612   0.216
## `Payment Typology 1`Medicaid                0.74536    0.12327   6.047
## `Payment Typology 1`Medicare                0.89221    0.13661   6.531
## `Payment Typology 1`Miscellaneous/Other     0.16509    0.40268   0.410
## `Payment Typology 1`Private Health Insurance  -0.33071    0.12027  -2.750
## `Payment Typology 1`Self-Pay               -1.41071    0.29338  -4.808
##                                            Pr(>|t|)
## (Intercept)                                 < 2e-16 ***
## `Age Group`18 to 29                         0.14513
## `Age Group`30 to 49                         0.00737 **
## `Age Group`50 to 69                         0.01001 *
## `Age Group`70 or Older                      0.29678
## GenderM                                    8.45e-06 ***
## `Type of Admission`Emergency                < 2e-16 ***
## `Type of Admission`Newborn                  0.00162 **
## `Type of Admission`Not Available            < 2e-16 ***
## `Type of Admission`Urgent                   < 2e-16 ***
## `Payment Typology 1`Federal/State/Local/VA  0.52914
## `Payment Typology 1`Managed Care, Unspecified  0.82923
## `Payment Typology 1`Medicaid               1.49e-09 ***
## `Payment Typology 1`Medicare               6.59e-11 ***
## `Payment Typology 1`Miscellaneous/Other     0.68182
## `Payment Typology 1`Private Health Insurance  0.00597 **
## `Payment Typology 1`Self-Pay               1.53e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 64.50696)
##
```

```
##     Null deviance: 4044001  on 58686  degrees of freedom
## Residual deviance: 3784623  on 58670  degrees of freedom
##   (35 observations deleted due to missingness)
## AIC: 411101
##
## 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     58677    3798620
## 2     58670    3784623  7    13996 < 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)	3.091	0.182	17.006	0.000	2.735	3.448
`Age Group`18 to 29	0.262	0.180	1.457	0.145	-0.091	0.615
`Age Group`30 to 49	0.432	0.161	2.680	0.007	0.116	0.747
`Age Group`50 to 69	0.410	0.159	2.575	0.010	0.098	0.722
`Age Group`70 or Older	0.193	0.185	1.043	0.297	-0.169	0.555
GenderM	0.306	0.069	4.454	0.000	0.172	0.441
`Type of Admission`Emergency	2.297	0.080	28.776	0.000	2.141	2.454
`Type of Admission`Newborn	0.565	0.179	3.152	0.002	0.214	0.917
`Type of Admission`Not Available	4.675	0.424	11.025	0.000	3.844	5.506
`Type of Admission`Urgent	5.864	0.112	52.148	0.000	5.644	6.084
`Payment Typology 1`Federal/State/Local/VA	-0.910	1.447	-0.629	0.529	-3.745	1.925
`Payment Typology 1`Managed Care, Unspecified	0.094	0.436	0.216	0.829	-0.761	0.949
`Payment Typology 1`Medicaid	0.745	0.123	6.047	0.000	0.504	0.987
`Payment Typology 1`Medicare	0.892	0.137	6.531	0.000	0.624	1.160
`Payment Typology 1`Miscellaneous/Other	0.165	0.403	0.410	0.682	-0.624	0.954
`Payment Typology 1`Private Health Insurance	-0.331	0.120	-2.750	0.006	-0.566	-0.095
`Payment Typology 1`Self-Pay	-1.411	0.293	-4.808	0.000	-1.986	-0.836