```
---
title: "HR ATTRITION DATA"
author: "David A. Bignotti"
date: "August 15, 2017"
output:
  word_document: default
  pdf_document: default
  html_document:
    keep_md: yes
---
```

```
```{}
#load libraries
library(ggplot2)
library(ggthemes)
library(dplyr)
library(survival)
library(survminer)
library(readr)
```
```

```
```{}
#load and clean data
dat <- read_csv("~/R/data/human-resources-analytics/HR_comma_sep.csv")
dat$salary<-ordered(data$salary,levels=c("low","medium","high"))
dat$Work_accident <- factor(data$Work_accident)
dat$promotion_last_5years <- factor(data$promotion_last_5years)
data <- dat %>% select(-sales)
str(data)
```
```

```
```{}
#create survival object
survival.object <- Surv(data$time_spend_company, data$left)
head(survival.object)
```
```

```
```{}
#print KM survival curve
fit.all <- survfit(survival.object~1)
summary(fit.all)
print(fit.all)
```
```

```
```{}
#primary survival curve
no1 <- ggsurvplot(fit.all,
           break.time.by = 1,
           palette = c("#E7B800", "#2E9FDF"),
           xlim = c(0,6),
           conf.int = TRUE, #add confidence interval
           pval = TRUE, #add p value
           risk.table = TRUE, #include risk table
           risk.table.height = 0.25,
           ggtheme = theme_light() # Change ggplot2 theme
)
```
```

```
```{no1, echo=FALSE}
#survival curves illustrated at the bottom of this document
no1
```
```

```{}
#stratify by salary range
fit.salary <- survfit(survival.object~data$salary)
print(fit.salary)
no2 <- ggsurvplot(fit.salary,
           break.time.by = 1,
           conf.int = TRUE,
           pval = TRUE,
           risk.table = TRUE,
           risk.table.col = "strata", #color by strata
           risk.table.height = 0.50,
           ggtheme = theme_light()
)
```

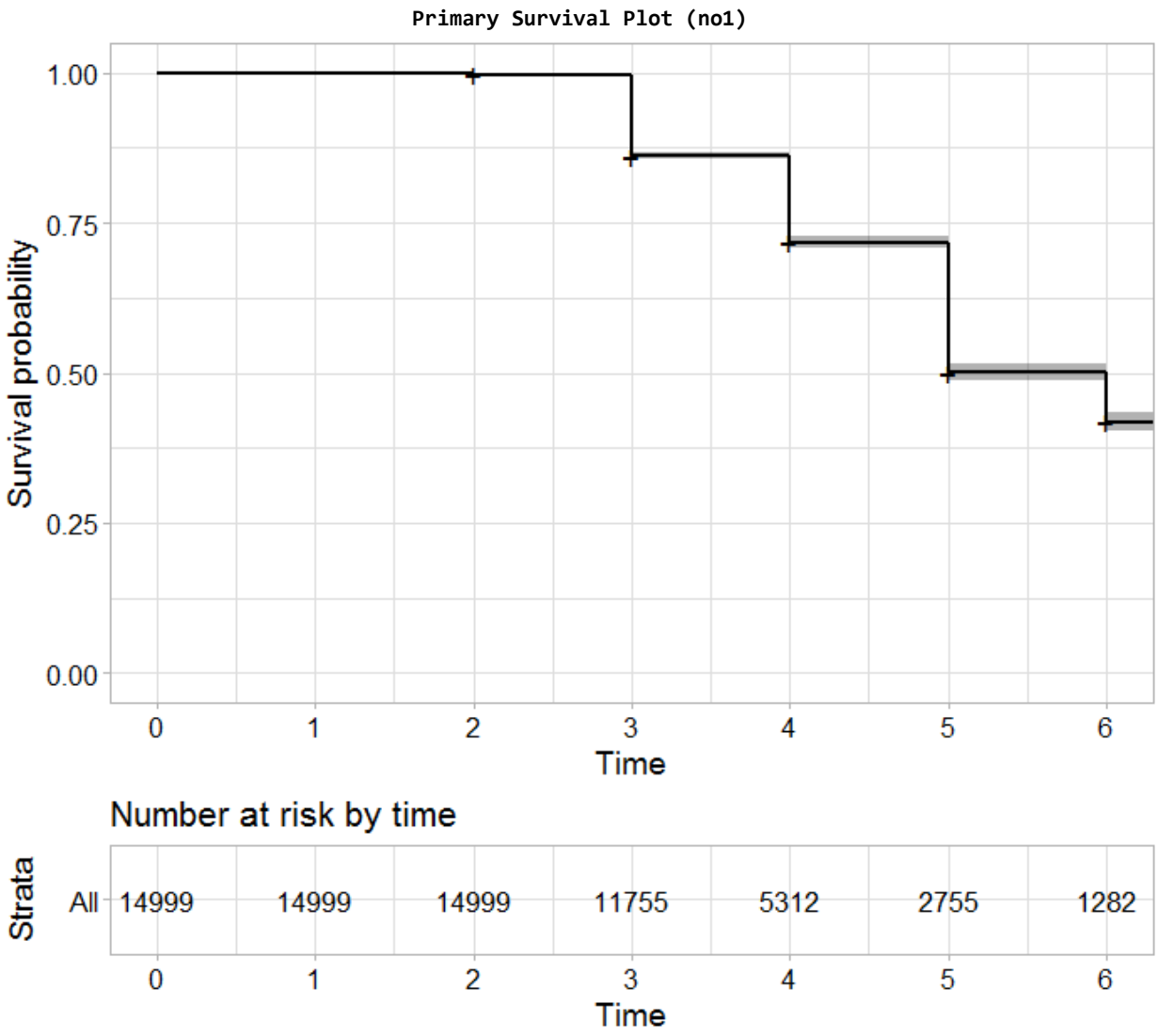
```{no2, echo=FALSE}
#survival curves illustrated at the bottom of this document
no2
```

```{}
#stratify by number of simultaneous projects
fit.projects <- survfit(survival.object~as.factor(data$number_project))
print(fit.projects)
no3 <- ggsurvplot(fit.projects,
           break.time.by = 1,
           conf.int = TRUE,
           pval = TRUE,
           risk.table = TRUE,
           risk.table.col = "strata",
           risk.table.height = 0.6,
           surv.plot.height = 0.9,
           ggtheme = theme_light()
)
```

```{no3, echo=FALSE}
#survival curves illustrated at the bottom of this document
no3
```

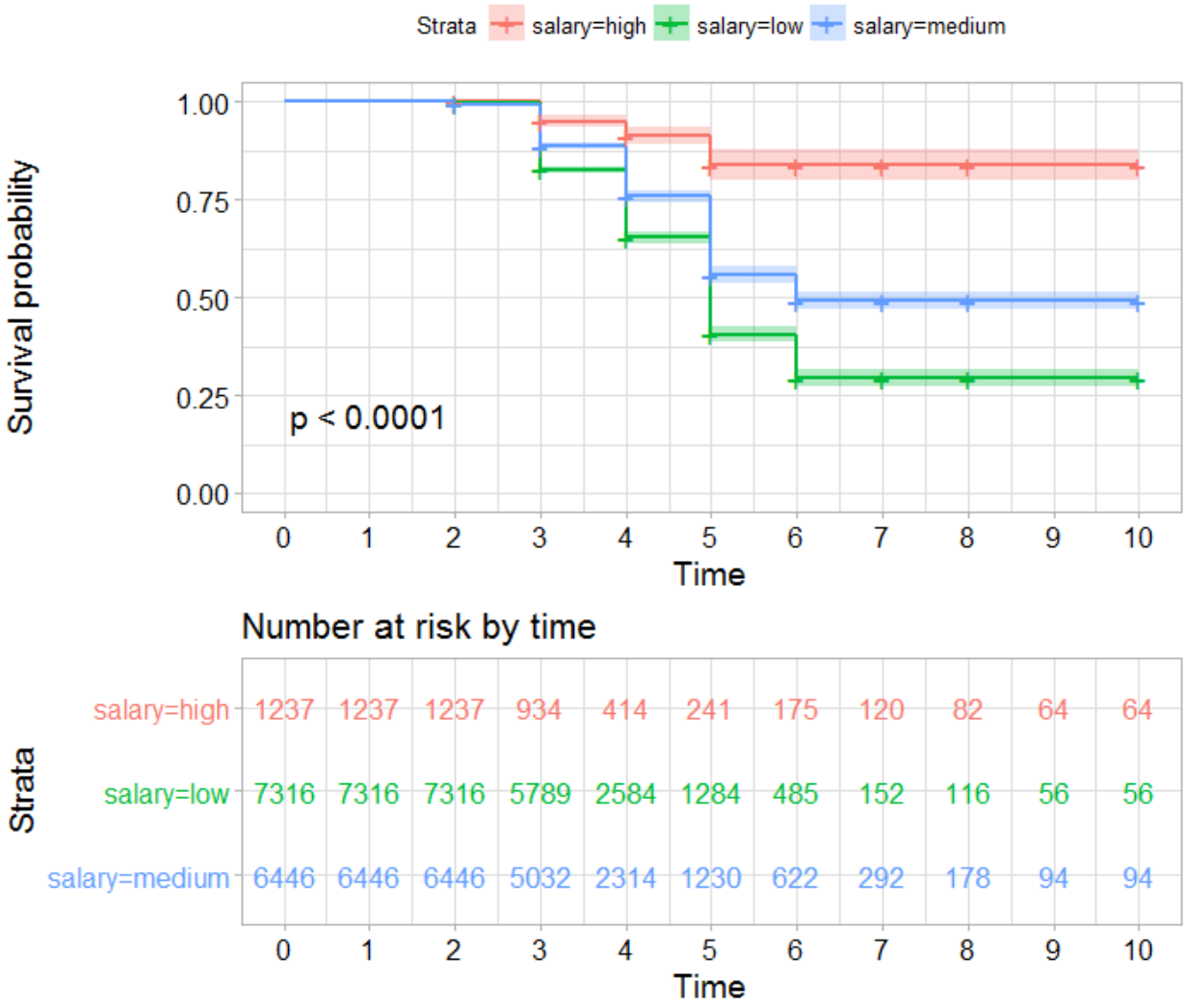
```{}
#promoted in the last five years
fit.promoted <- survfit(survival.object~as.factor(data$promotion_last_5years))
print(fit.promoted)
no4 <- ggsurvplot(fit.promoted,
           break.time.by = 1,
           conf.int = TRUE,
           pval = TRUE,
           risk.table = TRUE,
           risk.table.col = "strata",
           risk.table.height = 0.3,
           ggtheme = theme_light()
)
```

```{no4, echo=FALSE}
#survival curves illustrated at the bottom of this document
no4
```

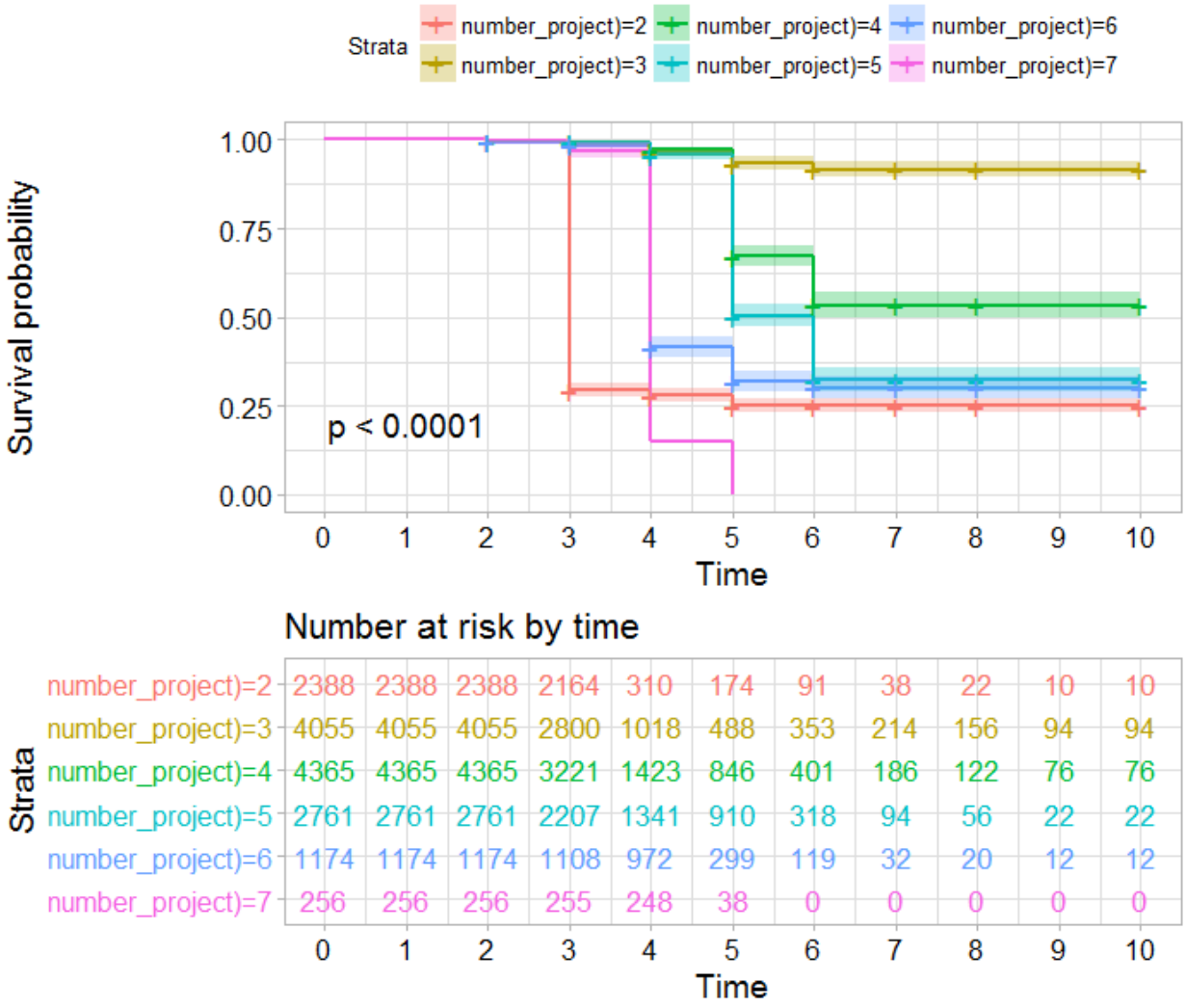

Primary Survival Plot (no1)
1.00
0.75
0.50
0.25
0.00
Survival probability
0
1
2
3
4
5
6
Time

Number at risk by time
Strata
All
14999
14999
14999
11755
5312
2755
1282
0
1
2
3
4
5
6
Time

Survival Plot Stratified by Salary Range (no2)



Survival Plot Stratified by Number of Concurrent Projects (no3)



Survival Plot Stratified by Yes/No Promotion Within the Last Five Years (no4)