

# L17 Survival

EPID 799B

Mike Dolan Fliss

Fall 2016

10.24.2016

# Survival (and other models)

- Quick review of lm/glm
  - Calls
  - predict() / confint() ☹️
  - Graphing
- Graphical exploration
- Packages
- Graphics
- Mice / missing patterns

# lm/glm

- Calls

- confint()
- coef()
- predict()

- ...or just go digging!

```
> model3.a1 = lm(data=births, preterm~mage)
> summary(model.a1)

Call:
glm(formula = preterm ~ mage, data = births)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.1213  -0.1146  -0.1120  -0.1089   0.8994

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.1265081   0.0053384   23.698  <2e-16 ***
mage        -0.0005181   0.0001896   -2.732   0.0063 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.09964519)

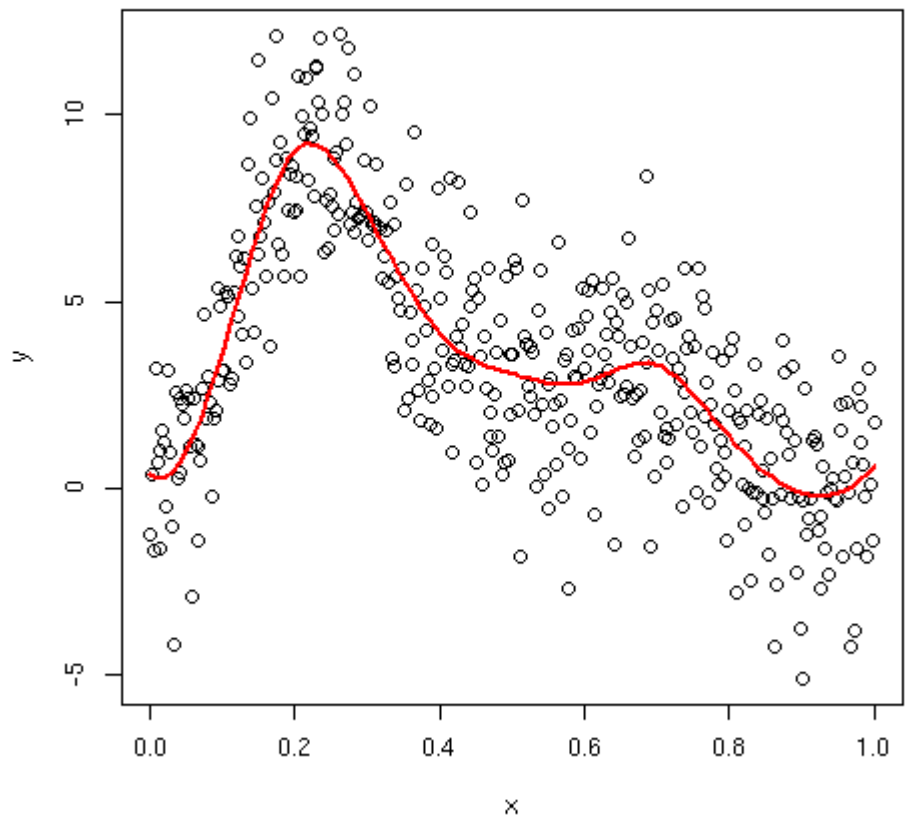
    Null deviance: 7797.1  on 78242  degrees of freedom
Residual deviance: 7796.3  on 78241  degrees of freedom
(7 observations deleted due to missingness)
AIC: 41609

Number of Fisher Scoring iterations: 2

> confint(model.a1)
Waiting for profiling to be done...
                2.5 %      97.5 %
(Intercept)  0.1160449795  0.1369711713
mage        -0.0008897407 -0.0001464102
> coef(model.a1)
(Intercept)      mage
0.1265080754 -0.0005180754
```

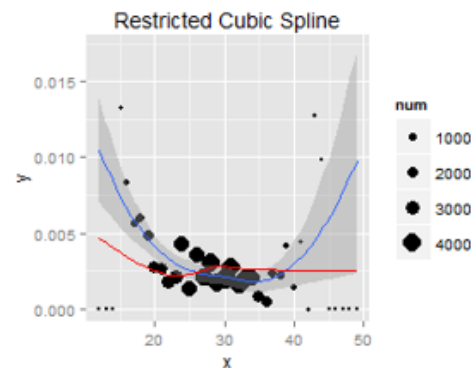
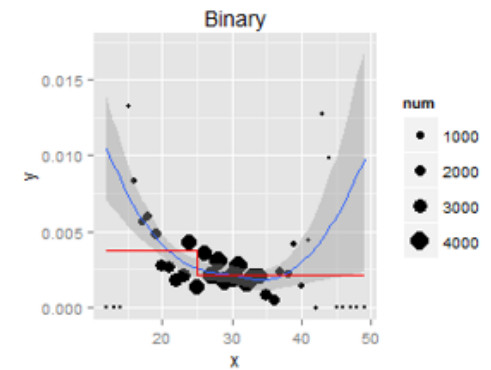
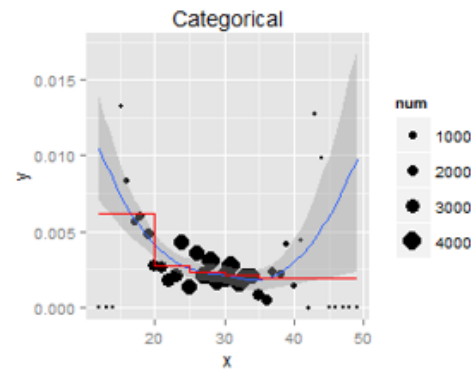
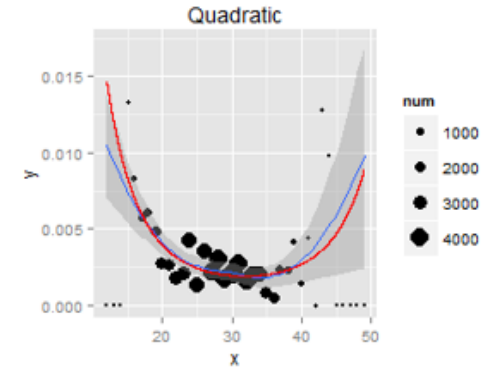
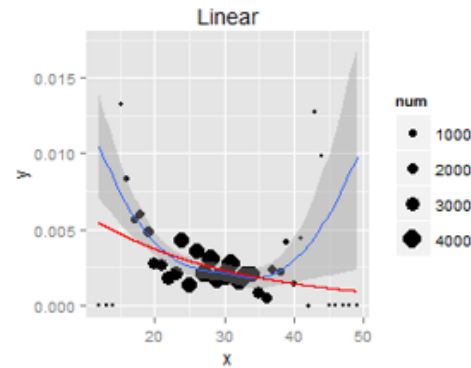
# lm/glm / functional form

- Graphical Exploration
  - Ggplot2 to the rescue: `Geom_smooth()`
    - `Method="lm"`
    - ...or "loess", etc.



# lm/glm / FF

- Function Form
  - R's got helpful spline functions.



# Consider other packages

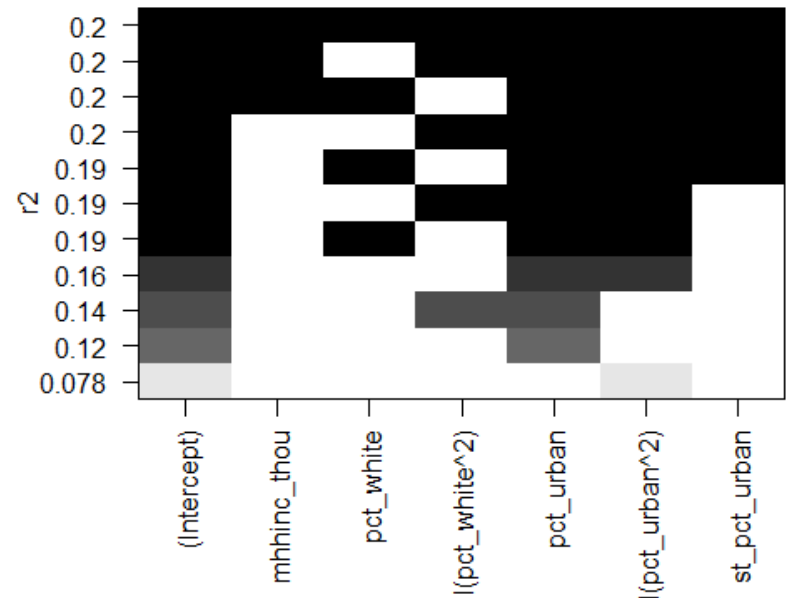
- E.g. MASS / leaps package. (Epi not allowed to do this. 😊)

```
Call:
lm(formula = trl_density ~ mhhinc_thou + poly(pct_white, 2) +
    poly(pct_urban, 2) + st_pct_urban, data = tobacco)

Residuals:
    Min       1Q   Median       3Q      Max
-1.7514 -0.3842 -0.1020  0.1969  4.1994

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.062246   0.226875   9.090  < 2e-16 ***
mhhinc_thou    -0.036294   0.023851  -1.522   0.128
poly(pct_white, 2)1  4.216238   0.767821   5.491 5.42e-08 ***
poly(pct_white, 2)2  0.414985   0.800530   0.518   0.604
poly(pct_urban, 2)1 -6.384823   0.844227  -7.563 1.12e-13 ***
poly(pct_urban, 2)2  5.118980   0.761918   6.719 3.55e-11 ***
st_pct_urban    -0.006617   0.002598  -2.547   0.011 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7282 on 775 degrees of freedom
Multiple R-squared:  0.1976,    Adjusted R-squared:  0.1914
F-statistic: 31.82 on 6 and 775 DF,  p-value: < 2.2e-16
```



```
library(MASS); library(leaps) #install.packages("leaps")
leaps=regsubsets(trl_density ~ mhhinc_thou + pct_white + l(pct_white^2)+
    pct_urban + l(pct_urban^2) + st_pct_urban, data=tobacco, nbest=2, nvmax=11, really.big=T)
regressum = summary(leaps)
plot(leaps, scale="r2") #selection plot, very helpful
```

# Consider other packages

- E.g. `broom::tidy()` up your results.

```
> summary(model.a1)

Call:
glm(formula = preterm ~ mage, data = births)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.1213  -0.1146  -0.1120  -0.1089   0.8994

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.1265081  0.0053384  23.698  <2e-16 ***
mage        -0.0005181  0.0001896  -2.732   0.0063 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

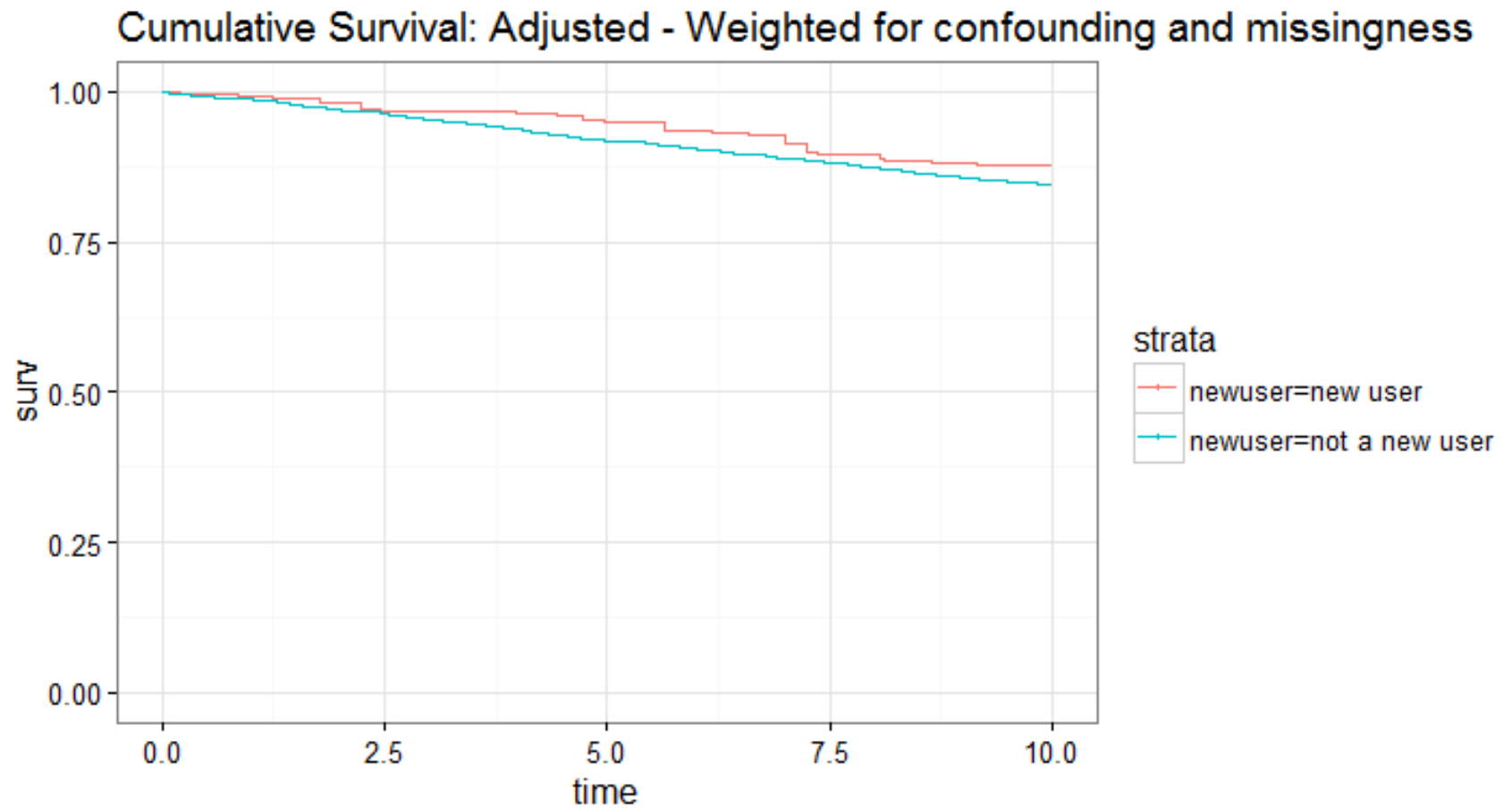
(Dispersion parameter for gaussian family taken to be 0.09964519)

Null deviance: 7797.1 on 78242 degrees of freedom
Residual deviance: 7796.3 on 78241 degrees of freedom
(7 observations deleted due to missingness)
AIC: 41609

Number of Fisher Scoring iterations: 1
```

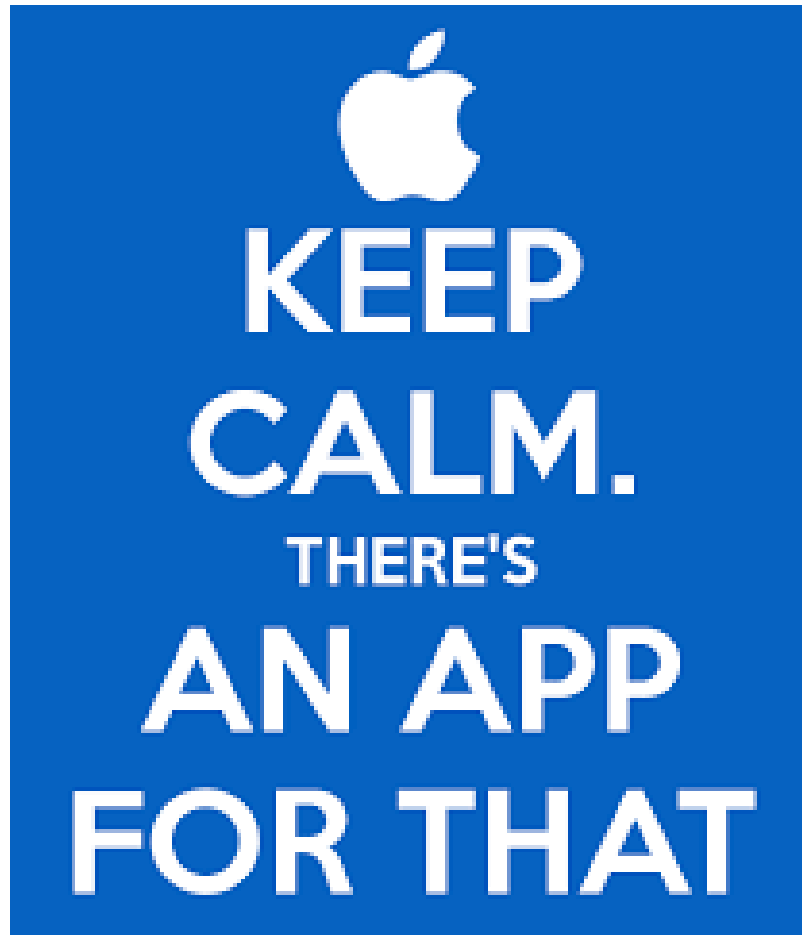
```
> tidy(model.a1)
# A tibble: 2 x 5
  term            estimate std.error statistic    p.value
  <fct>            <dbl>    <dbl>     <dbl>  <dbl>
1 (Intercept)  0.1265080754  0.0053384123  23.697697 1.041943e-123
2 mage      -0.0005180754  0.0001896286  -2.732053  6.295509e-03
```

# Survival





# Survival



# Survival



# Survival



# Survival

- The ... you guessed it... survival package!
- Also:

```
m.crude = coxph(Surv(t, event)~newuser, data=d, ties="efron")
m.saturated = coxph(Surv(t, event)~newuser+age+male+white+obese+dbp+sbp+diabetes+hyplipid+htn+smoke, data=d, ties="efron")
summary(m.saturated)
m.adj = coxph(Surv(t, event)~newuser+age+obese+diabetes+hyplipid+htn+smoke, data=d, ties="efron") #leave out bp?
m.adj2 = coxph(Surv(t, event)~newuser+age+dbp+sbp+diabetes+hyplipid+htn, data=d, ties="efron") #leave out bp?
m.adj3 = coxph(Surv(t, event)~newuser+age+diabetes+hyplipid+htn, data=d, ties="efron") #leave out bp?
m.w1 = coxph(Surv(t, event)~newuser, data=d[d$bpmissing=="BP measured",], weights = w.iptw, ties="efron") #leave out bp?
m.w2 = coxph(Surv(t, event)~newuser, data=d[d$bpmissing=="BP measured",], weights = w.missing, ties="efron")
m.w3 = coxph(Surv(t, event)~newuser+cluster(patcode), data=d.split, weights = w.censoring, ties="efron")
m.wall = coxph(Surv(t, event)~newuser+cluster(patcode), data=d.split, weights = w.all, ties="efron")

all.m = list(m.1=m.crude, m.2=m.saturated, m.3=m.adj, m.4=m.adj2, m.5=m.adj3,
             m.6=m.w1, m.7=m.w2, m.8 = m.w3, m.9 = m.wall)
```

# Survival

- The ... you guessed it... survival package!
- Also other helper packages, like:
  - ggAlly, survutils, ipw packages...

# Survival

- Let's jump right in.

```
d.split = survSplit(Surv(t, event)~., data=d,  
  cut=as.numeric(dropout.quantiles), end="t", start="time0", event="event")
```

```
m.crude = coxph(Surv(t, event)~newuser, data=d, ties="efron")  
m.saturated = coxph(Surv(t, event)~newuser+age+male+white+obese+dbp+sbp+diabetes+hyplipid+htn+smoke, data=d, ties="efron")  
summary(m.saturated)  
m.adj = coxph(Surv(t, event)~newuser+age+obese+diabetes+hyplipid+htn+smoke, data=d, ties="efron") #leave out bp?  
m.adj2 = coxph(Surv(t, event)~newuser+age+dbp+sbp+diabetes+hyplipid+htn, data=d, ties="efron") #leave out bp?  
m.adj3 = coxph(Surv(t, event)~newuser+age+diabetes+hyplipid+htn, data=d, ties="efron") #leave out bp?  
m.w1 = coxph(Surv(t, event)~newuser, data=d[d$bpmissing=="BP measured",], weights = w.iptw, ties="efron") #leave out  
m.w2 = coxph(Surv(t, event)~newuser, data=d[d$bpmissing=="BP measured",], weights = w.missing, ties="efron")  
m.w3 = coxph(Surv(t, event)~newuser+cluster(patcode), data=d.split, weights = w.censoring, ties="efron")  
m.wall = coxph(Surv(t, event)~newuser+cluster(patcode), data=d.split, weights = w.all, ties="efron")  
  
all.m = list(m.1=m.crude, m.2=m.saturated, m.3=m.adj, m.4=m.adj2, m.5=m.adj3,  
  m.6=m.w1, m.7=m.w2, m.8 = m.w3, m.9 = m.wall)
```

# Survival

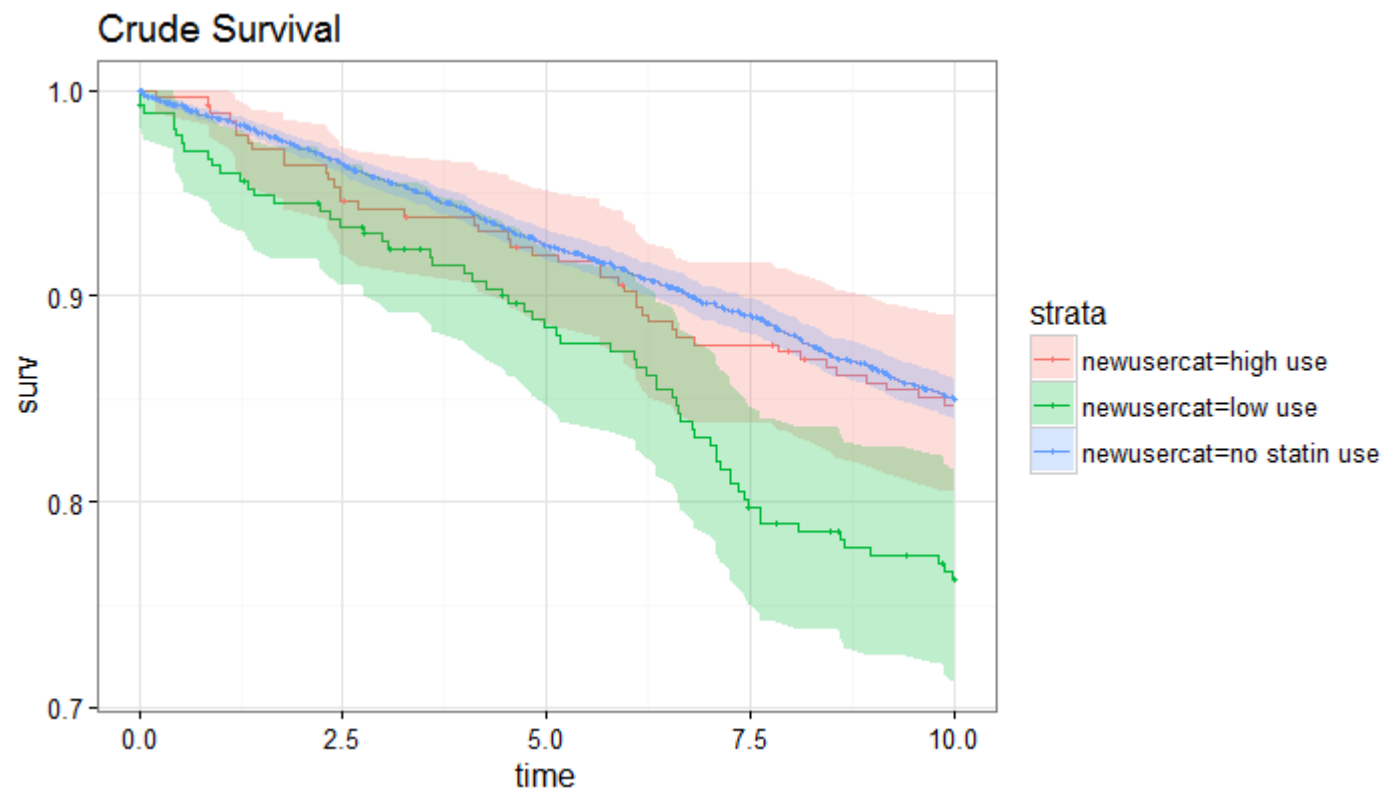
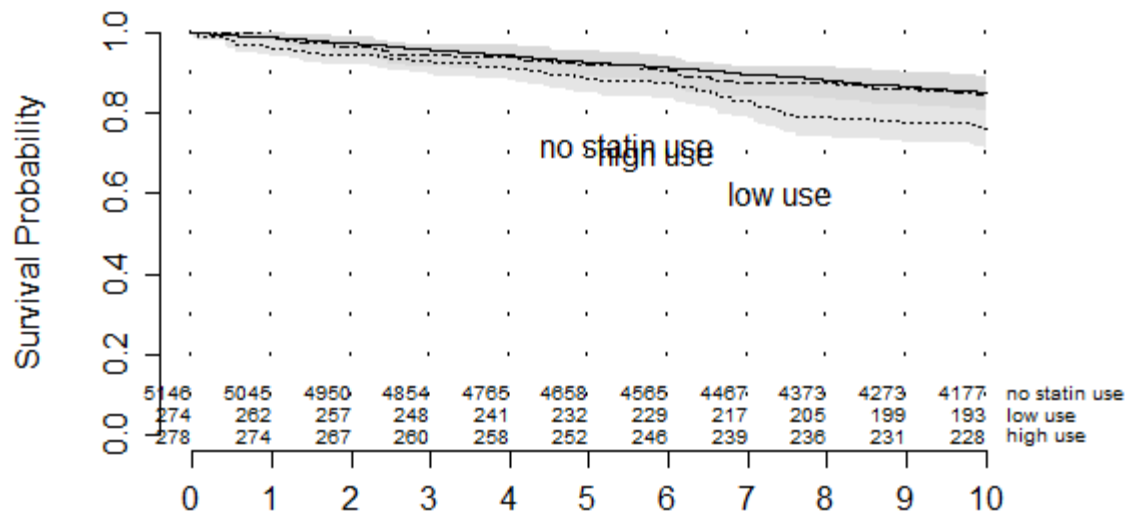
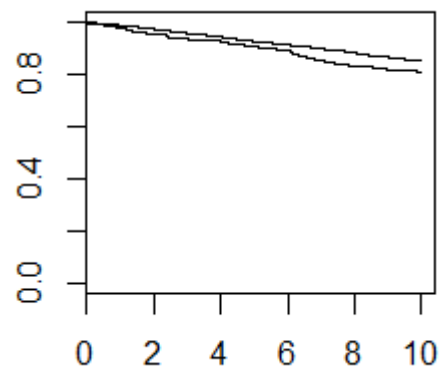
- Plots

```
#shell.exec("http://rstudio-pubs-static.s3.amazonaws.com/5588_72eb65bfbe0a4cb7b655d2eee0751584.html") #Look at hazard
#http://rstudio-pubs-static.s3.amazonaws.com/16003_8d6f2069af094998bfbb3d3a7dfb1c17.html
```

```
d$survival = Surv(d$t, d$event==1)
km.bycat = survfit(survival~exposure, data=d)
plot(km.bycat) #base R. Meh.
survplot(npsurv(formula = Surv(t, event==1) ~ newuser, data=d), xlab="Years of Follow-up",
         levels.only = T, n.risk = T,dots=T)
survplot(npsurv(formula = Surv(t, event==1) ~ newusercat, data=d), xlab="Years of Follow-up",
         levels.only = T, n.risk = T,dots=T)
```

```
#even better, ggplot2 themed plots referenced below. Also consider ggally package:
# http://www.r-statistics.com/2013/07/creating-good-looking-survival-curves-the-ggsurv-function/
# great intro to ggplot: http://timchurches.github.io/ggplot2er/#1
# functionhttp://www.ceb-institute.org/bbs/wp-content/uploads/2011/09/handout\_ggplot2.pdf
```

```
#Triple stratification : just to know about it. NOTE High use looks a lot like no use...?!
t.survfit <- survfit(Surv(t, event)~newusercat, data=d, type="kaplan-meier"); plot(t.survfit)
t.survframe <- createSurvivalFrame(t.survfit)
ggplot(data = t.survframe, aes(colour = strata, group = strata)) +
  geom_step(aes(x = time, y = surv), direction = "hv") +
  geom_ribbon(aes(x = time, ymax = upper, ymin = lower, fill = strata), linetype = 0, alpha = 0.25) +
  geom_point(data = subset(t.survframe, n.censor == 1), aes(x = time, y = surv), shape = 20) + theme_bw() +
  ggtitle("Crude Survival")
```

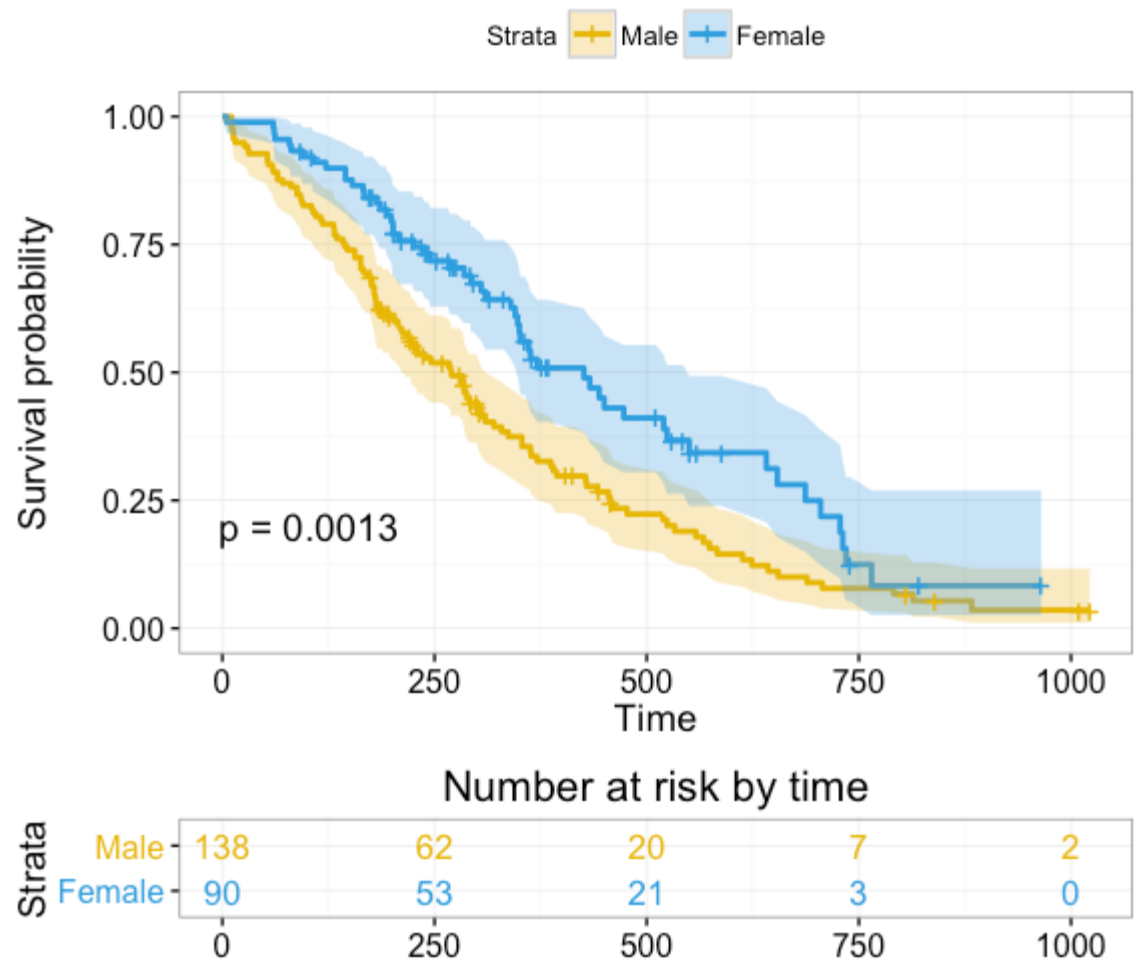




# Survival

- survminer

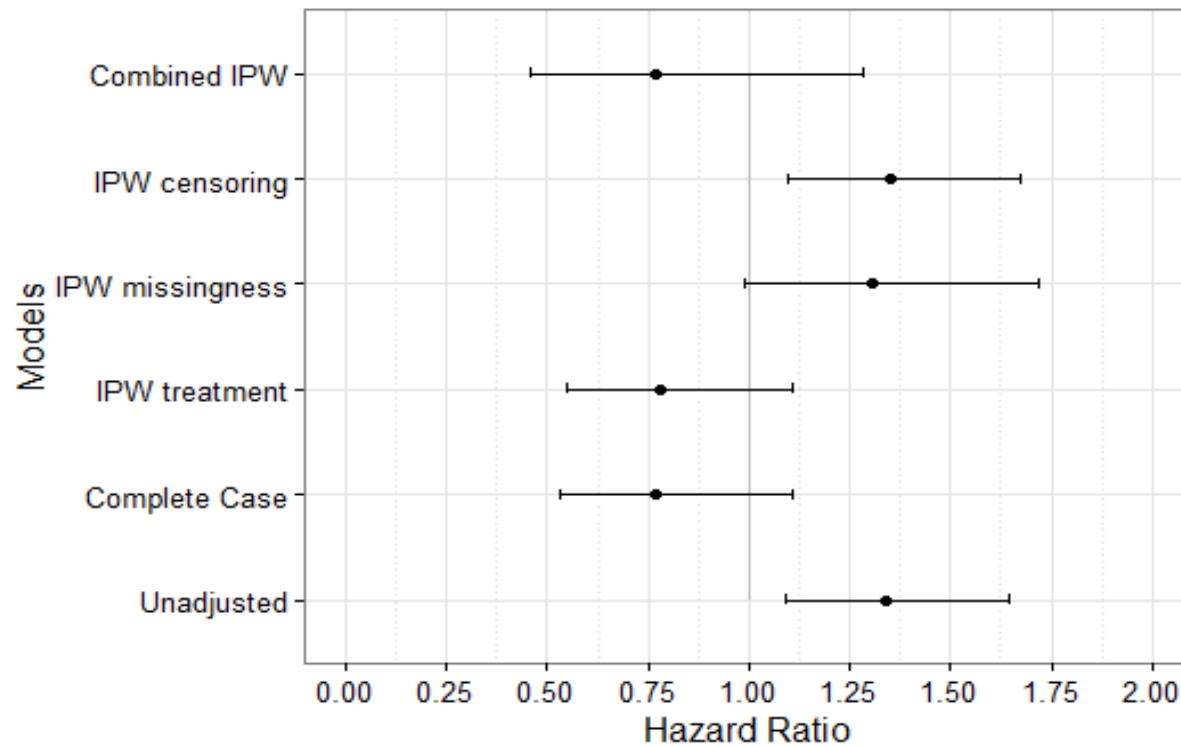
<http://www.sthda.com/english/rpkgs/survminer/>



```
ggsurvplot(fit, size = 1, # change line size
  palette = c("#E7B800", "#2E9FDF"), # custom color palettes
  conf.int = TRUE, # Add confidence interval
  pval = TRUE, # Add p-value
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Risk table color by groups
  legend.labs = c("Male", "Female"), # Change legend labels
  risk.table.height = 0.25, # Useful to change when you have multiple groups
  ggtheme = theme_bw() # Change ggplot2 theme
)
```

# Survival

- Model Outputs



**Figure 1.** Estimated Effect of Initiation of Moderate- or High-Dose Statin Therapy on 10-year Cardiovascular Disease Endpoints in 5,689 Adults, National Ambulatory Care Medical Survey, 2005-2009. See Table 2 for model parameters. Abbreviations: HR, hazard ratio; IPW, inverse probability weighted

**Table 2.** Estimated Effect of Initiation of Moderate- or High-Dose Statin Therapy on 10-year Cardiovascular Disease Endpoints in 5,689 Adults, National Ambulatory Care Medical Survey, 2005-2009

# Survival

- Tables!

Model and Statin Therapy	No. of events	Person-years of follow-up	HR	95% CI
Unadjusted				
No Statin Therapy	447	46,608	1.	
Statin Therapy	105	4,839	1.34	1.09, 1.64
Complete case <sup>a</sup> multivariable outcome model				
No Statin Therapy	441	23,720	1.	
Statin Therapy	97	4,460	0.77	0.53, 1.11
IPW treatment <sup>b</sup> model				
No Statin Therapy	729	44,259	1.	
Statin Therapy	59	4,708	0.78	0.55, 1.11
IPW missingness <sup>c</sup> model				
No Statin Therapy	757	46,560	1.	
Statin Therapy	97	5,179	1.30	0.99, 1.72
IPW censoring <sup>d</sup> model				
No Statin Therapy	755	46,591	1.	
Statin Therapy	105	4,853	1.35	1.10, 1.67 <sup>e</sup>
Combined IPW model				
No Statin Therapy	417	24,765	1.	
Statin Therapy	22	1,689	0.77	0.46, 1.28 <sup>e</sup>

Abbreviations: HR, hazard ratio; IPW, inverse probability weighted; BP, blood pressure; M/H, Moderate or High

<sup>a</sup> Complete case analysis does not include blood pressure measurements, missing in 45% of the records, and adjusted for age, sex, race, systolic and diastolic blood pressure, diabetes, hyperlipidemia, hypertension and smoking status.

<sup>b</sup> Treatment model adjusted for age, sex, race, obesity, diabetes, systolic and diastolic blood pressure, hyperlipidemia, hypertension and smoking status.

<sup>c</sup> BP missingness model adjusted for age, sex, race, obesity, diabetes, hyperlipidemia, hypertension and smoking status.

<sup>d</sup> IPW censoring model adjusted for sex, race, obesity, hypertension and smoking status.

<sup>e</sup> Robust 95% CI

# Survival: R is always growing

- <https://www.r-statistics.com/2013/07/creating-good-looking-survival-curves-the-ggsurv-function/>
  - <https://ggobi.github.io/ggally/docs.html#ggallyggsurv>
  - ^ NOTE: Now incorporated into GGally::ggsurv(). Nice!
- <https://github.com/tinyheero/survutils>
- Shout outs to Ann Von Holle for helping me in 722 with survival in R.