

## Appendix

### Appendix 1: cross tables

	0	1	Total
has_name			
FALSE	7,984 (66%)	4,141 (34%)	12,125 (100%)
TRUE	8,811 (23%)	29,036 (77%)	37,847 (100%)
<b>Total</b>	16,795 (34%)	33,177 (66%)	49,972 (100%)

	0	1	Total
sex			
Female	8,162 (34%)	15,767 (66%)	23,929 (100%)
Male	8,633 (33%)	17,410 (67%)	26,043 (100%)
<b>Total</b>	16,795 (34%)	33,177 (66%)	49,972 (100%)

	0	1	Total
spay_neuter			
Fixed	7,146 (19%)	29,865 (81%)	37,011 (100%)
Intact	9,649 (74%)	3,312 (26%)	12,961 (100%)
<b>Total</b>	16,795 (34%)	33,177 (66%)	49,972 (100%)

	0	1	Total
animal_type			
Other	346 (64%)	197 (36%)	543 (100%)
Cat	8,799 (47%)	9,962 (53%)	18,761 (100%)
Dog	7,650 (25%)	23,018 (75%)	30,668 (100%)
<b>Total</b>	16,795 (34%)	33,177 (66%)	49,972 (100%)

	0	1	Total
color_count			
1	8,462 (39%)	13,254 (61%)	21,716 (100%)
2	7,972 (30%)	18,894 (70%)	26,866 (100%)
3	361 (26%)	1,029 (74%)	1,390 (100%)
<b>Total</b>	16,795 (34%)	33,177 (66%)	49,972 (100%)

	0	1	Total
	0	1	Total
<b>color_shade_intens</b>			
Unknown	1,018 (46%)	1,175 (54%)	2,193 (100%)
Dark	3,033 (36%)	5,441 (64%)	8,474 (100%)
Light	7,726 (31%)	17,236 (69%)	24,962 (100%)
Medium	5,018 (35%)	9,325 (65%)	14,343 (100%)
<b>Total</b>	16,795 (34%)	33,177 (66%)	49,972 (100%)

Cross tables details for Unknown sex (removed from the final dataset), after removing the negative values for our variable `time_to_outcome` (-5).

	Female	Male	Unknown	Total
<b>outcome</b>				
0	8,162 (34%)	8,633 (33%)	4,335 (98%)	21,130 (39%)
1	15,767 (66%)	17,410 (67%)	96 (2%)	33,273 (61%)
<b>Total</b>	23,929 (100%)	26,043 (100%)	4,431 (100%)	54,403 (100%)

	Female	Male	Unknown	Total
<b>animal_type</b>				
Cat	9,476 (40%)	9,285 (36%)	1,796 (41%)	20,557 (38%)
Dog	14,227 (59%)	16,441 (63%)	161 (4%)	30,829 (57%)
Other	226 (1%)	317 (1%)	2,474 (56%)	3,017 (6%)
<b>Total</b>	23,929 (100%)	26,043 (100%)	4,431 (100%)	54,403 (100%)

	Female	Male	Unknown	Total
<b>has_name</b>				
FALSE	5,961 (25%)	6,164 (24%)	4,306 (97%)	16,431 (30%)
TRUE	17,968 (75%)	19,879 (76%)	125 (3%)	37,972 (70%)
<b>Total</b>	23,929 (100%)	26,043 (100%)	4,431 (100%)	54,403 (100%)

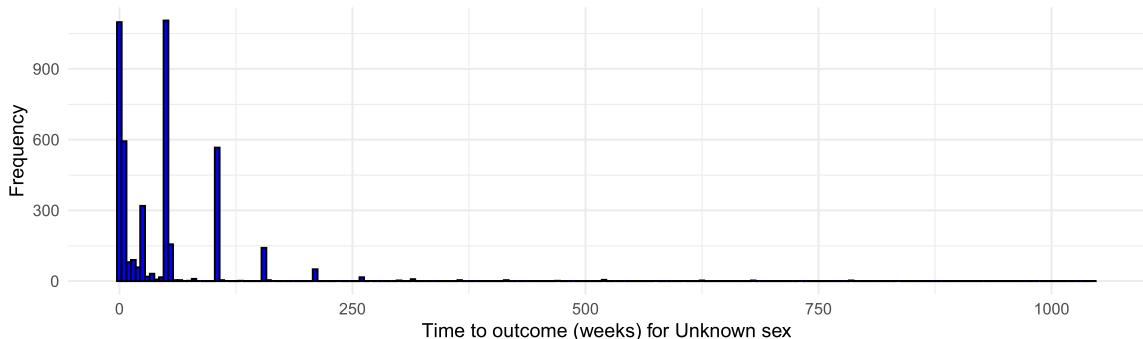
	Female	Male	Unknown	Total
<b>spay_neuter</b>				

	Female	Male	Unknown	<b>Total</b>
Fixed	17,594 (74%)	19,417 (75%)	0 (0%)	37,011 (68%)
Intact	6,335 (26%)	6,626 (25%)	0 (0%)	12,961 (24%)
Unknown	0 (0%)	0 (0%)	4,431 (100%)	4,431 (8%)
<b>Total</b>	23,929 (100%)	26,043 (100%)	4,431 (100%)	54,403 (100%)

	Female	Male	Unknown	<b>Total</b>
<b>color_count</b>				
1	10,962 (46%)	10,754 (41%)	2,782 (63%)	24,498 (45%)
2	12,326 (52%)	14,540 (56%)	1,625 (37%)	28,491 (52%)
3	641 (3%)	749 (3%)	24 (1%)	1,414 (3%)
<b>Total</b>	23,929 (100%)	26,043 (100%)	4,431 (100%)	54,403 (100%)

	Female	Male	Unknown	<b>Total</b>
<b>color_shade_intens</b>				
Dark	4,278 (18%)	4,196 (16%)	652 (15%)	9,126 (17%)
Light	11,449 (48%)	13,513 (52%)	1,101 (25%)	26,063 (48%)
Medium	6,375 (27%)	7,968 (31%)	2,579 (58%)	16,922 (31%)
Unknown	1,827 (8%)	366 (1%)	99 (2%)	2,292 (4%)
<b>Total</b>	23,929 (100%)	26,043 (100%)	4,431 (100%)	54,403 (100%)

Histogram of Values of time\_to\_outcome in weeks for Unknown sex



## **Appendix 2: Nonparametric methods**

### **KM and FH methods**

#### **Kaplan-Meier and Fleming-harrington estimators**

\*\*\* Kaplan-Meier estimator \*\*\*

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ 1, data = data[, -1])
```

	n	events	median	0.95LCL	0.95UCL
[1,]	49972	33177	106	106	106

\*\*\* Fleming-harrington estimator \*\*\*

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ 1, data = data[, -1], type = "fleming-harrington")
```

	n	events	median	0.95LCL	0.95UCL
[1,]	49972	33177	106	106	106

\*\*\* Estimated probability of not being adopted within 1 year old \*\*\*

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ 1, data = data[, -1])
```

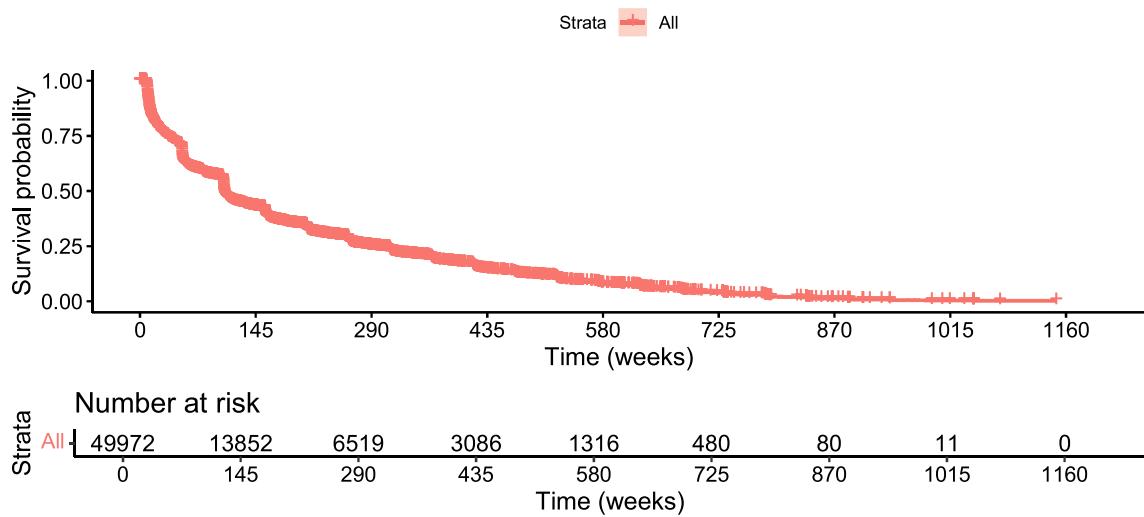
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
52	29309	12989	0.708	0.00217	0.703		0.712	

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ 1, data = data[, -1], type = "fleming-harrington")
```

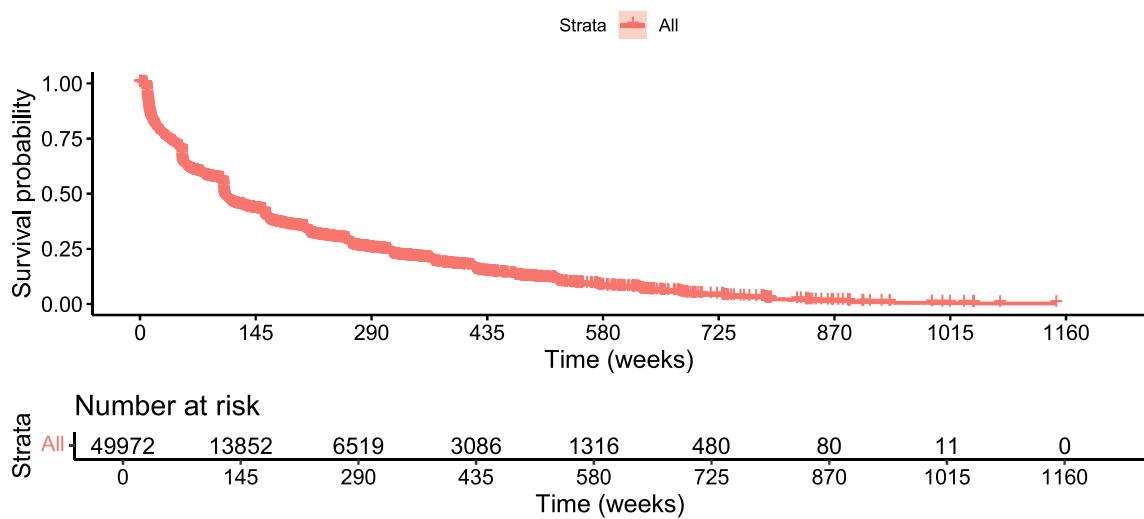
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
52	29309	12989	0.708	0.00217	0.704		0.712	

\*\*\* Kaplan-Meier and Fleming-Harrington estimators \*\*\*

### Kaplan–Meier estimator



### Fleming–harrington estimator

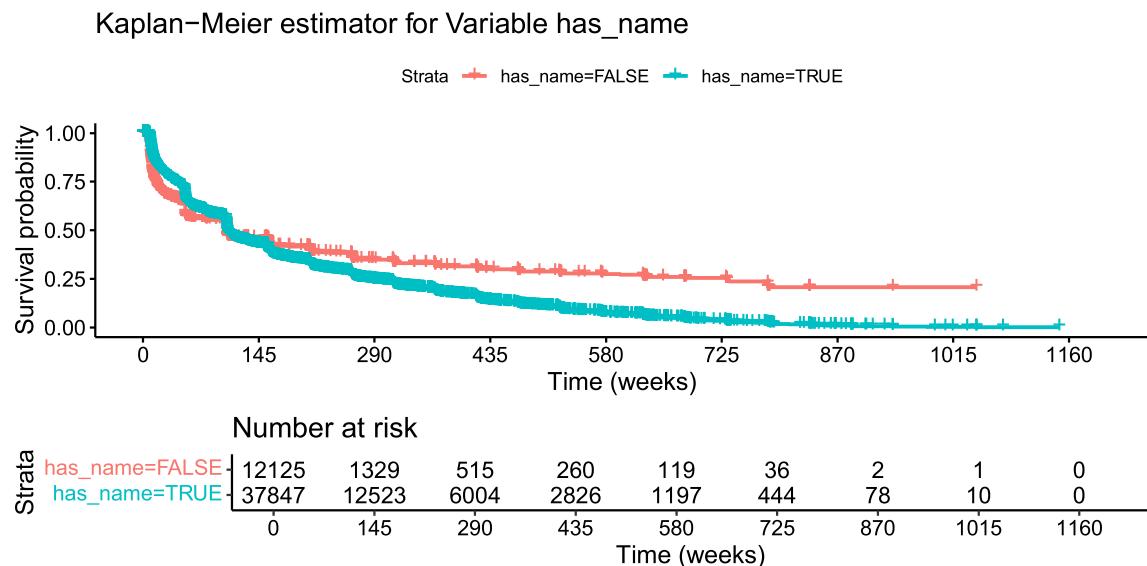


### comparison of groups

#### Variable has\_name

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ has_name,
  data = data[, -1])
```

	n	events	median	0.95LCL	0.95UCL
has_name=FALSE	12125	4141	105	105	105
has_name=TRUE	37847	29036	106	106	107



Call:

```
survdiff(formula = Surv(time_to_outcome, outcome) ~ has_name,
          data = data[, -1])
```

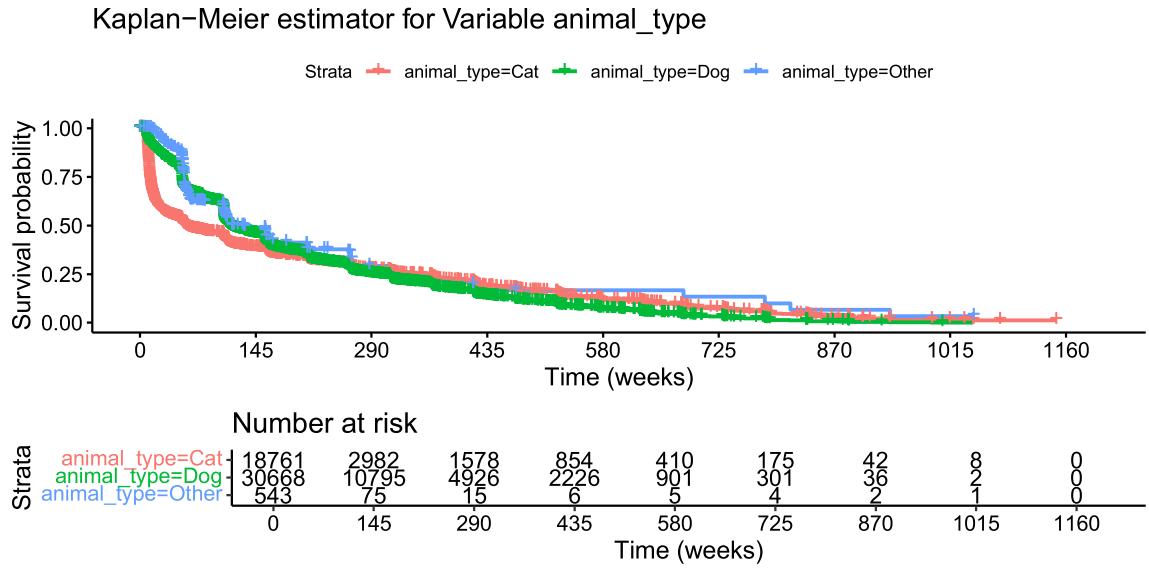
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
has_name=FALSE	12125	4141	4172	0.2329	0.272
has_name=TRUE	37847	29036	29005	0.0335	0.272

Chisq= 0.3 on 1 degrees of freedom, p= 0.6

### Variable animal\_type

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ animal_type,
              data = data[, -1])
```

	n	events	median	0.95LCL	0.95UCL
animal_type=Other	543	197	112.6	107.1	163.0
animal_type=Cat	18761	9962	56.9	54.7	60.3
animal_type=Dog	30668	23018	111.6	110.0	113.9



Call:

```
survdiff(formula = Surv(time_to_outcome, outcome) ~ animal_type,
          data = data)
```

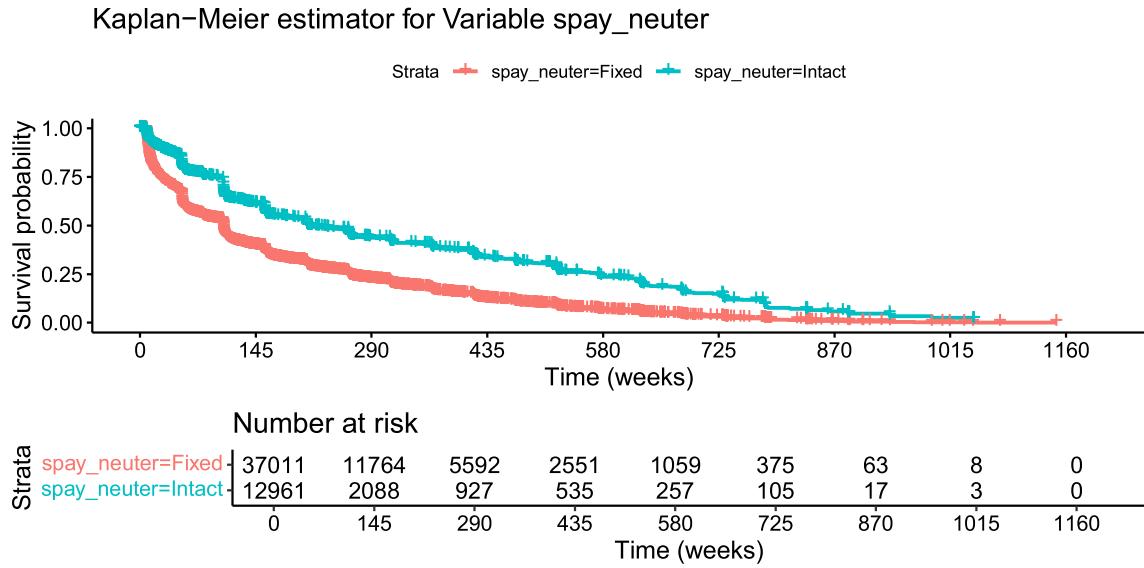
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
animal_type=Other	543	197	295	32.6	33.2
animal_type=Cat	18761	9962	8437	275.7	375.8
animal_type=Dog	30668	23018	24445	83.3	322.0

Chisq= 398 on 2 degrees of freedom, p= <2e-16

### Variable spay\_neuter

Call: survfit(formula = Surv(time\_to\_outcome, outcome) ~ spay\_neuter,  
data = data[, -1])

	n	events	median	0.95LCL	0.95UCL
spay_neuter=Fixed	37011	29865	105	105	105
spay_neuter=Intact	12961	3312	210	209	243



Call:

```
survdiff(formula = Surv(time_to_outcome, outcome) ~ spay_neuter,
          data = data[, -1])
```

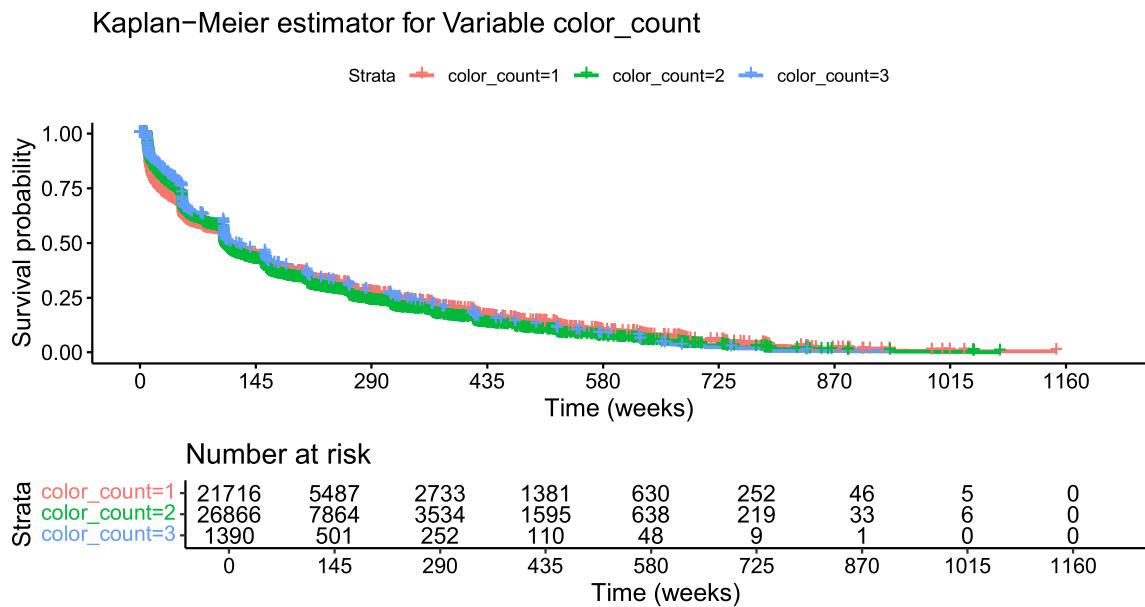
	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
spay_neuter=Fixed	37011	29865	27385	225	1297
spay_neuter=Intact	12961	3312	5792	1062	1297

Chisq= 1297 on 1 degrees of freedom, p= <2e-16

Variable color\_count

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ color_count,
              data = data[, -1])
```

	n	events	median	0.95LCL	0.95UCL
color_count=1	21716	13254	106	105	106
color_count=2	26866	18894	106	106	106
color_count=3	1390	1029	112	106	139



Call:

```
survdiff(formula = Surv(time_to_outcome, outcome) ~ color_count,
          data = data[, -1])
```

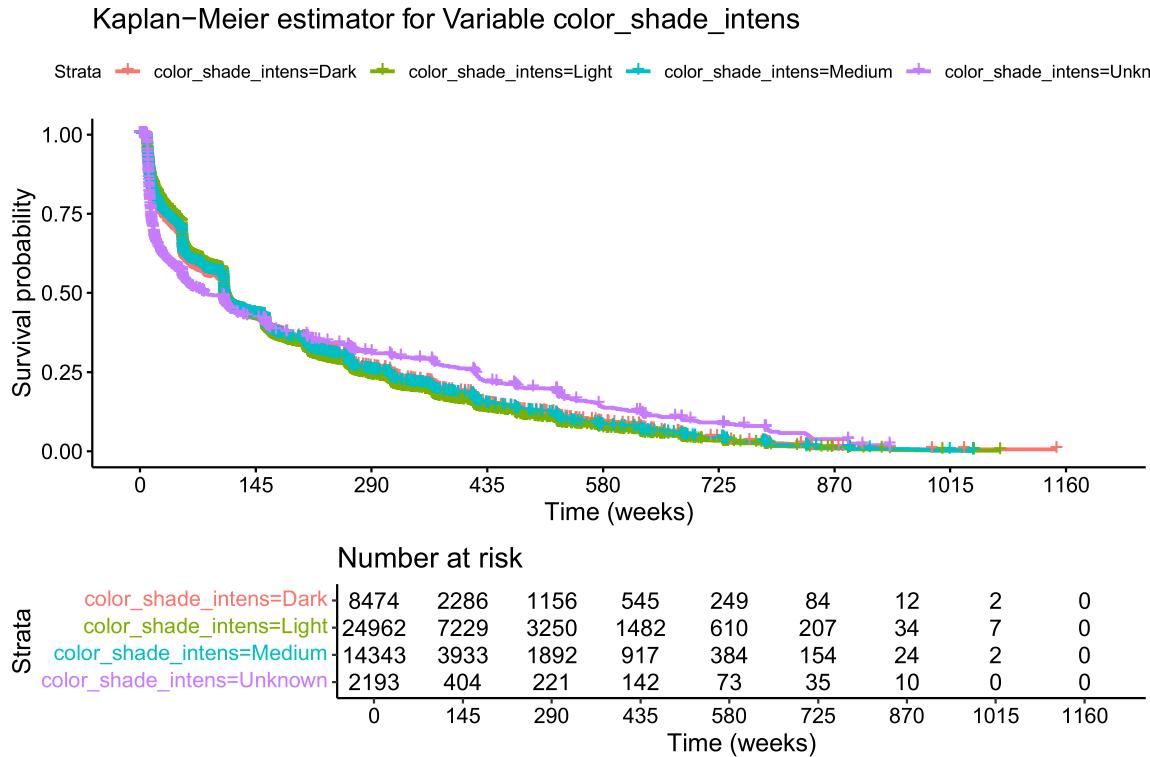
	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
color_count=1	21716	13254	13501	4.52	7.69
color_count=2	26866	18894	18579	5.33	12.21
color_count=3	1390	1029	1097	4.17	4.34

Chisq= 14.1 on 2 degrees of freedom, p= 9e-04

Variable color\_shade\_intens

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ color_shade_intens,
              data = data[, -1])
```

	n	events	median	0.95LCL	0.95UCL
color_shade_intens=Unknown	2193	1175	79.9	62	106
color_shade_intens=Dark	8474	5441	105.4	105	106
color_shade_intens=Light	24962	17236	106.1	106	107
color_shade_intens=Medium	14343	9325	105.6	105	106



Call:

```
survdiff(formula = Surv(time_to_outcome, outcome) ~ color_shade_intens,
          data = data[, -1])
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
color_shade_intens=Unknown	2193	1175	1141	0.991	1.034
color_shade_intens=Dark	8474	5441	5486	0.373	0.450
color_shade_intens=Light	24962	17236	17153	0.405	0.843
color_shade_intens=Medium	14343	9325	9397	0.547	0.767

Chisq= 2.3 on 3 degrees of freedom, p= 0.5

### Stratification and comparison of more than 2 groups

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ sex + strata(spay_neuter),
              data = data)
```

	n	events	median	0.95LCL	0.95UCL
sex=Female, strata(spay_neuter)=Fixed	17594	14348	104	104	105

sex=Female, strata(spay_neuter)=Intact	6335	1419	264	261	314
sex=Male, strata(spay_neuter)=Fixed	19417	15517	105	105	105
sex=Male, strata(spay_neuter)=Intact	6626	1893	203	170	209

Call:

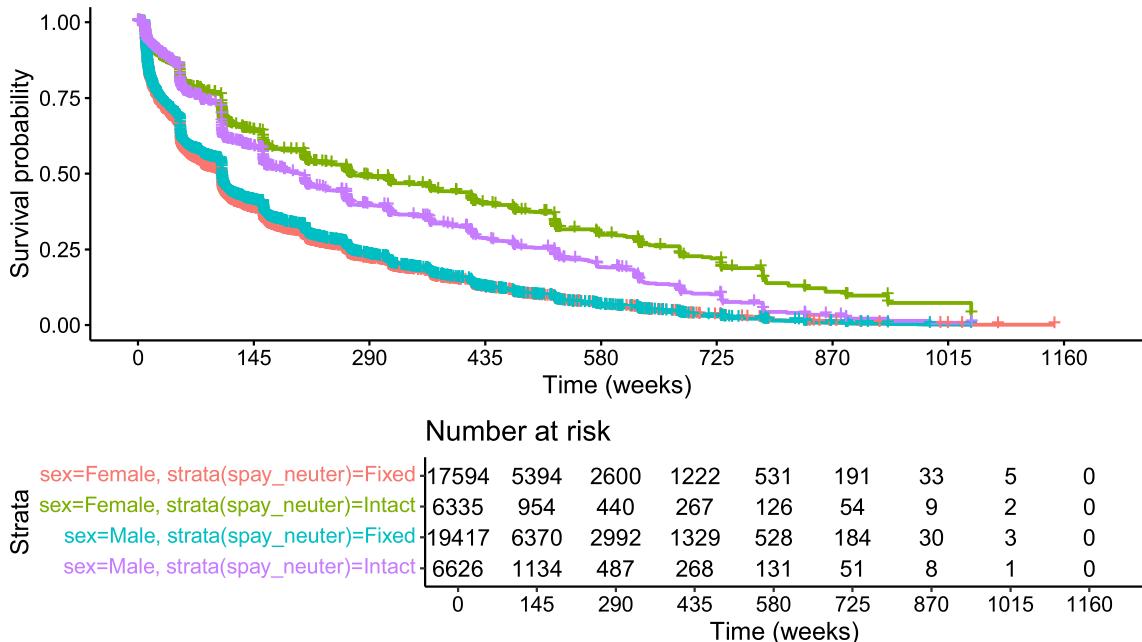
```
survdiff(formula = Surv(time_to_outcome, outcome) ~ sex + strata(spay_neuter),
          data = data)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
sex=Female	23929	15767	15587	2.07	3.93
sex=Male	26043	17410	17590	1.84	3.93

Chisq= 3.9 on 1 degrees of freedom, p= 0.05

### Kaplan–Meier estimator for compounding groups sex + spay\_neuter

=Female, strata(spay\_neuter)=Fixed    sex=Female, strata(spay\_neuter)=Intact    sex=Male, strata(spay\_neuter)=Fixed    sex=Male, s



### Appendix 3: Semi-parametric multivariate Cox regression

Start: AIC=645865.6

Surv(time\_to\_outcome, outcome) ~ animal\_type + has\_name + sex +

```
spay_neuter + color_count + color_shade_intens
```

	Df	AIC
<none>		645866
- sex	1	645866
- color_shade_intens	3	645877
- has_name	1	645896
- color_count	2	645911
- animal_type	2	646210
- spay_neuter	1	647312

### Result of the automatic model selection.

Call:

```
coxph(formula = Surv(time_to_outcome, outcome) ~ animal_type +
    has_name + sex + spay_neuter + color_count + color_shade_intens,
    data = data)
```

n= 49972, number of events= 33177

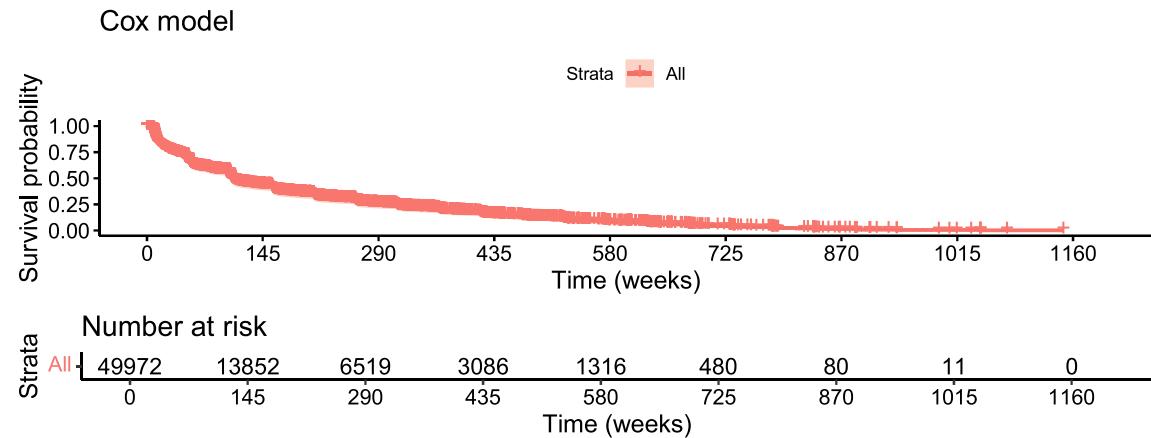
	coef	exp(coef)	se(coef)	z	Pr(> z )	
animal_typeCat	0.23545	1.26548	0.07279	3.235	0.001217	**
animal_typeDog	-0.01486	0.98525	0.07232	-0.205	0.837231	
has_nameTRUE	-0.10005	0.90479	0.01749	-5.720	1.07e-08	***
sexMale	-0.01872	0.98145	0.01110	-1.688	0.091487	.
spay_neuterIntact	-0.66741	0.51303	0.01899	-35.153	< 2e-16	***
color_count2	0.10392	1.10951	0.01488	6.985	2.84e-12	***
color_count3	0.02813	1.02853	0.03509	0.802	0.422733	
color_shade_intensDark	0.12966	1.13844	0.03378	3.839	0.000124	***
color_shade_intensLight	0.09097	1.09524	0.03374	2.696	0.007015	**
color_shade_intensMedium	0.11213	1.11866	0.03271	3.428	0.000607	***
---						
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'
	0.05	'.'	0.1	' '	1	

	exp(coef)	exp(-coef)	lower .95	upper .95
animal_typeCat	1.2655	0.7902	1.0972	1.4595
animal_typeDog	0.9853	1.0150	0.8550	1.1353
has_nameTRUE	0.9048	1.1052	0.8743	0.9363
sexMale	0.9814	1.0189	0.9603	1.0030
spay_neuterIntact	0.5130	1.9492	0.4943	0.5325
color_count2	1.1095	0.9013	1.0776	1.1423
color_count3	1.0285	0.9723	0.9602	1.1018

color_shade_intensDark	1.1384	0.8784	1.0655	1.2164
color_shade_intensLight	1.0952	0.9130	1.0251	1.1701
color_shade_intensMedium	1.1187	0.8939	1.0492	1.1927

Concordance= 0.615 (se = 0.002 )  
 Likelihood ratio test= 1936 on 10 df, p=<2e-16  
 Wald test = 1707 on 10 df, p=<2e-16  
 Score (logrank) test = 1752 on 10 df, p=<2e-16

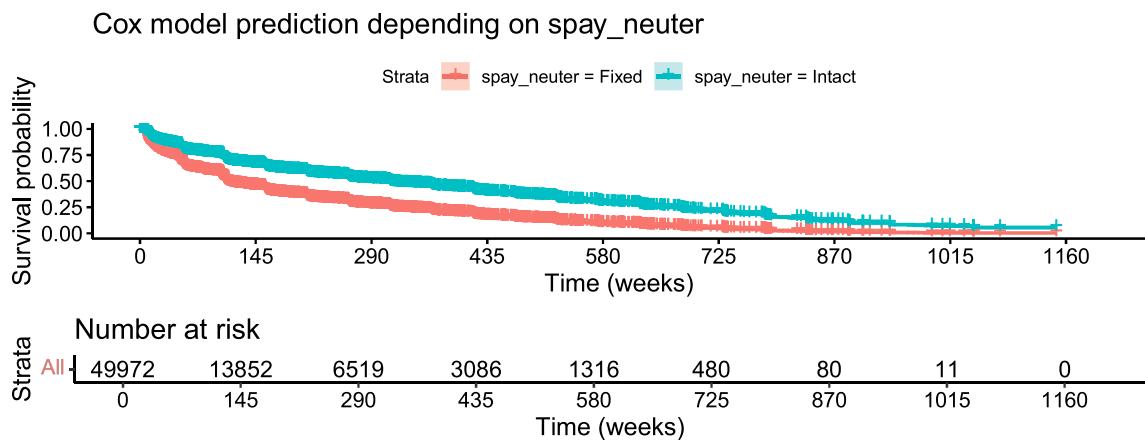
### Probability survival prediction



```
Call: survfit(formula = cox_model, data = data_df)
```

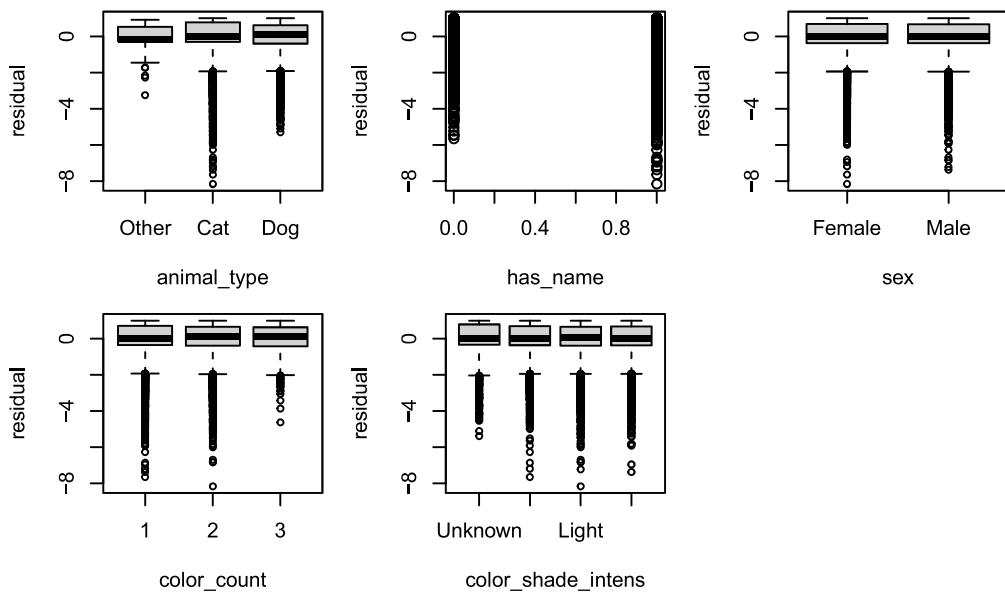
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
52	29309	12989	0.714	0.0193	0.677		0.752	

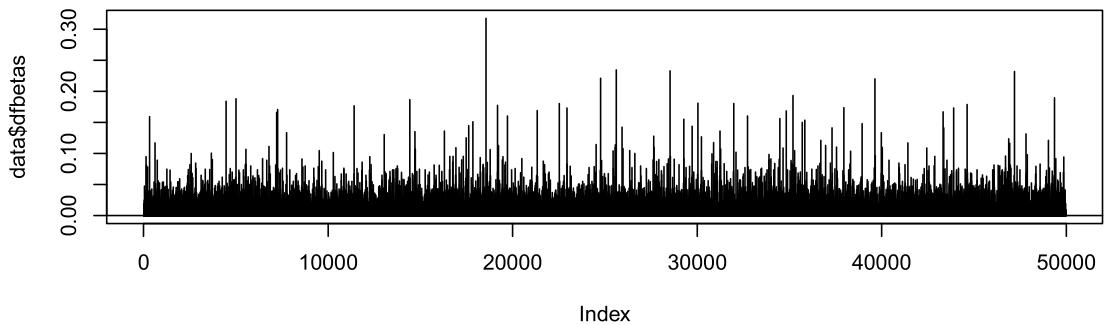
New data point prediction for one point changing only the spay\_neuter variable



#### Appendix 4: Model validation

##### Model validation check with Martingale Residuals



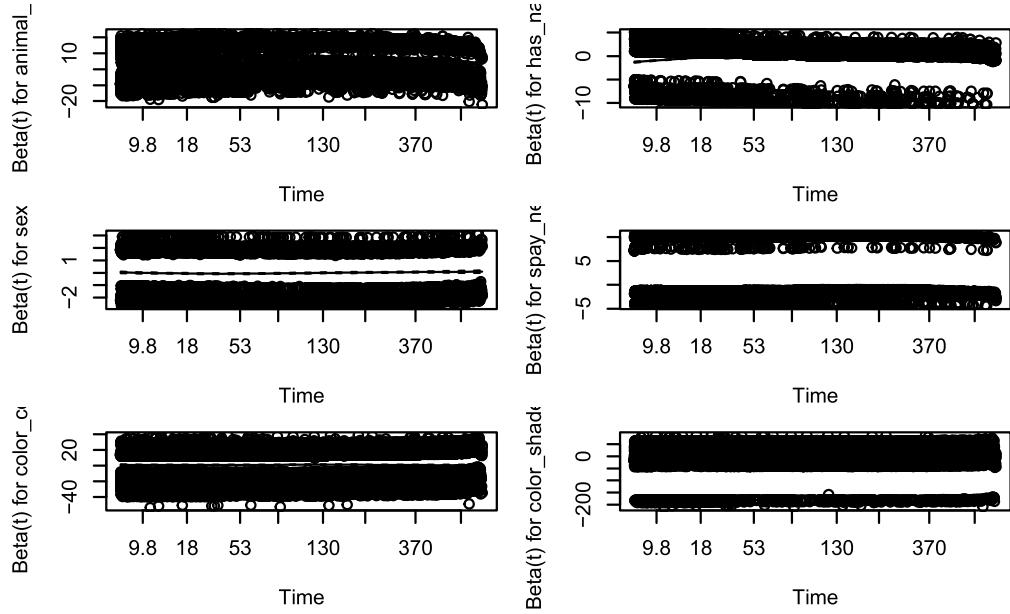


We could try to identify data with the highest values, remove them and redo the all process to eventually check if the model assumptions are satisfied.

### **Model validation check with Proportionality of hazards**

Schoenfeld residuals

	chisq	df	p
animal_type	3125.4	2	< 2e-16
has_name	986.9	1	< 2e-16
sex	24.3	1	8.2e-07
spay_neuter	29.8	1	4.7e-08
color_count	258.6	2	< 2e-16
color_shade_intens	297.1	3	< 2e-16
GLOBAL	3789.7	10	< 2e-16



The figure indicates violation of proportionality of the hazards.

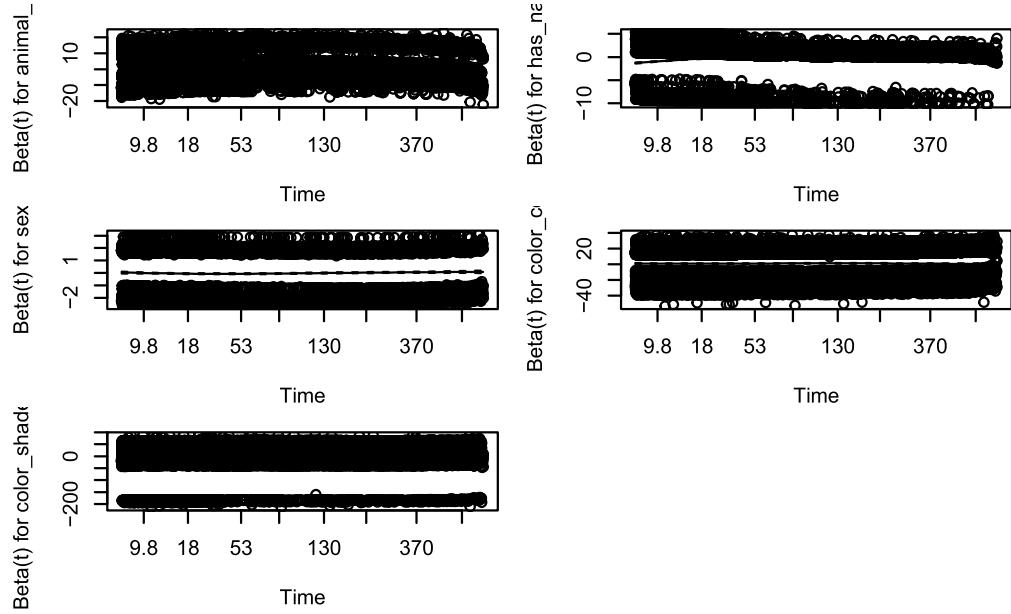
It is not better by trying to fix it with stratification.

```
fit <- coxph(Surv(time_to_outcome, outcome) ~ animal_type + has_name + sex + color_count + color_shade_intens)

residual.sch <- cox.zph(fit)
residual.sch
```

	chisq	df	p
animal_type	3115.8	2	< 2e-16
has_name	1110.6	1	< 2e-16
sex	25.1	1	5.5e-07
color_count	258.0	2	< 2e-16
color_shade_intens	298.3	3	< 2e-16
GLOBAL	3781.4	9	< 2e-16

```
par(mfrow = c(3, 2), mar = c(4, 4, 1, 1), oma = c(0, 0, 2, 0))
plot(residual.sch)
```



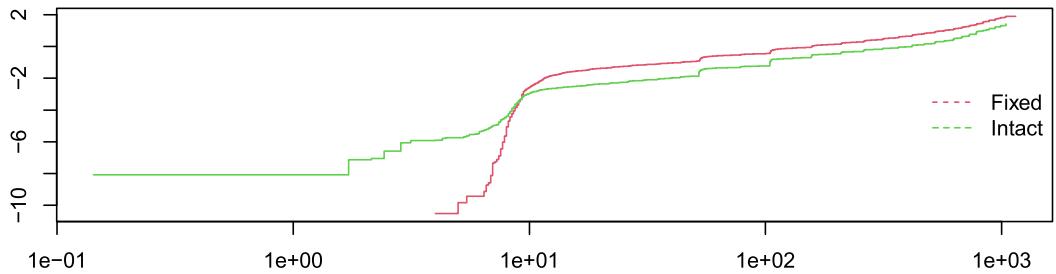
**Let's look at KM model as well.**

```
fit.KM<-survfit(Surv(time_to_outcome, outcome)~spay_neuter, data=data)
fit.KM
```

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ spay_neuter,
  data = data)
```

	n	events	median	0.95LCL	0.95UCL
spay_neuter=Fixed	37011	29865	105	105	105
spay_neuter=Intact	12961	3312	210	209	243

```
plot(fit.KM, fun ="cloglog", col=2:3)
legend("right", legend=sort(unique(data$spay_neuter)), col=2:3, lty=2:3, horiz=FALSE, bty='n')
```



The figure indicates violation of proportionality of the hazards.

Let's try to fix it with stratification

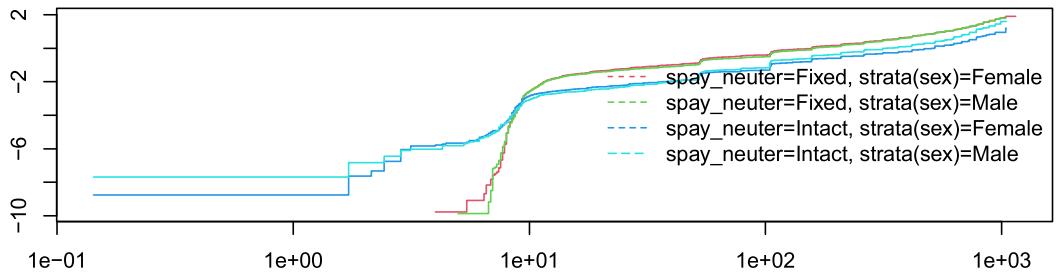
```
fit.KM_strata<-survfit(Surv(time_to_outcome, outcome)~spay_neuter+strata(sex),data=data)
fit.KM_strata
```

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ spay_neuter +
  strata(sex), data = data)
```

	n	events	median	0.95LCL	0.95UCL
spay_neuter=Fixed, strata(sex)=Female	17594	14348	104	104	105
spay_neuter=Fixed, strata(sex)=Male	19417	15517	105	105	105
spay_neuter=Intact, strata(sex)=Female	6335	1419	264	261	314
spay_neuter=Intact, strata(sex)=Male	6626	1893	203	170	209

```
plot(fit.KM_strata, fun ="cloglog", col=2:5)
```

```
legend("right", legend=sort(names(fit.KM_strata$strata))), col=2:5, lty=2:5, horiz=FALSE, bt
```



The figure still indicates violation of proportionality of the hazards.

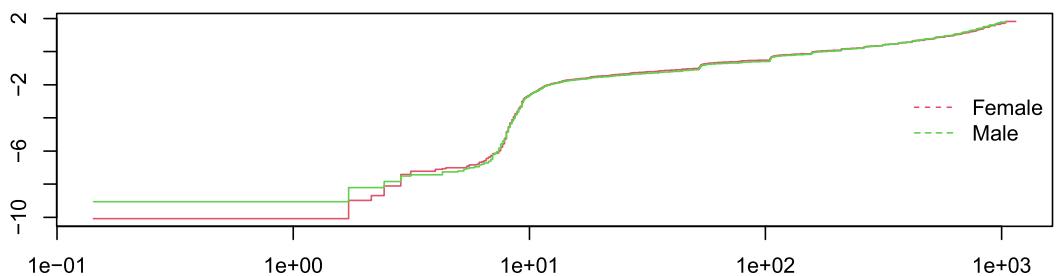
Let's try by using sex first.

```
fit.KM<-survfit(Surv(time_to_outcome, outcome)~sex, data=data)
fit.KM
```

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ sex, data = data)
```

	n	events	median	0.95LCL	0.95UCL
sex=Female	23929	15767	105	105	106
sex=Male	26043	17410	106	106	107

```
plot(fit.KM, fun ="cloglog", col=2:3)
legend("right", legend=sort(unique(data$sex)), col=2:3, lty=2:3, horiz=FALSE, bty='n')
```



The figure indicates violation of proportionality of the hazards.

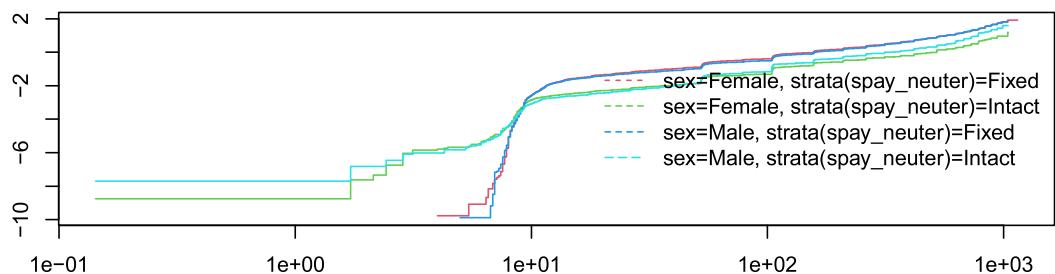
Let's try to fix it with stratification

```
fit.KM_strata<-survfit(Surv(time_to_outcome, outcome)~sex+strata(spay_neuter), data=data)
fit.KM_strata
```

```
Call: survfit(formula = Surv(time_to_outcome, outcome) ~ sex + strata(spay_neuter),
  data = data)
```

	n	events	median	0.95LCL	0.95UCL
sex=Female, strata(spay_neuter)=Fixed	17594	14348	104	104	105
sex=Female, strata(spay_neuter)=Intact	6335	1419	264	261	314
sex=Male, strata(spay_neuter)=Fixed	19417	15517	105	105	105
sex=Male, strata(spay_neuter)=Intact	6626	1893	203	170	209

```
plot(fit.KM_strata, fun ="cloglog", col=2:5)
legend("right", legend=sort(names(fit.KM_strata$strata)), col=2:5, lty=2:5, horiz=FALSE, bty="n")
```



The figure still indicates violation of proportionality of the hazards.