Fitting Topic Numbers

```
df <- data.frame(p = c(5,15,20,26),n=c(952,3260,8254,21937))
df$k <- c(50,65,90,120)

m1 <- lm(k~n+p,df)
m2 <- lm(k~n,df)
m3 <- lm(k~poly(n,2),df)
m4 <- lm(k~log(n),df)
```

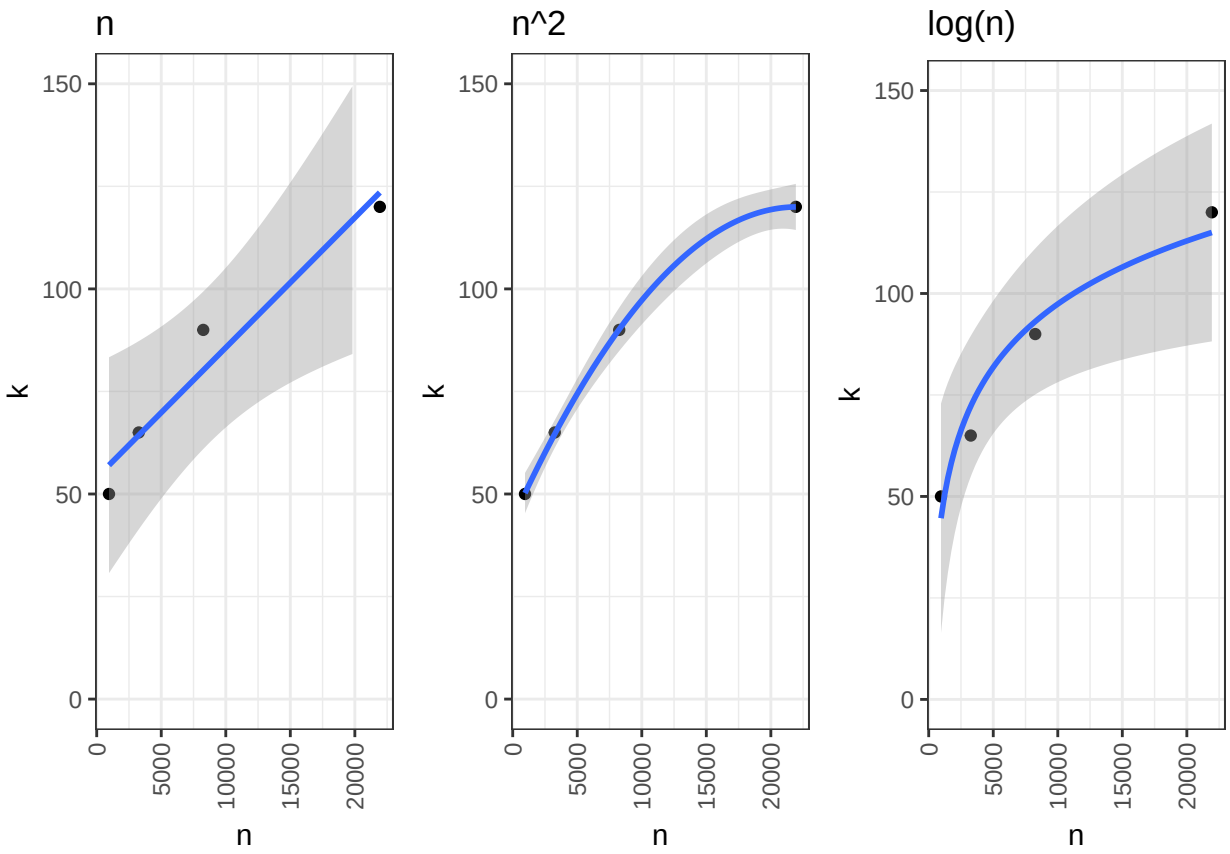
```
p <- ggplot(df, aes(n,k)) +
  geom_point() +
  ylim(0,150) +
  theme_bw() +
  theme(
    axis.text.x = element_text(angle = 90, hjust = 1, vjust=0.5)
  )


p1 <- p + stat_smooth(method="lm") + ggtitle("n")

p2 <- p + stat_smooth(method="lm", formula = y ~ poly(x,2)) +
  ggtitle("n^2")

p3 <- p + stat_smooth(method="lm", formula = y ~ log(x)) +
  ggtitle("log(n)")

multiplot(p1, p2, p3, cols=3)
```




```
stargazer(m1,m2,m3,m4, header=FALSE)
```

Table 1:

	Dependent variable:			
	k			
	(1)	(2)	(3)	(4)
n	0.002	0.003**		
	(0.001)	(0.001)		
p	1.558			
	(0.796)			
poly(n, 2)1			51.573***	
			(0.442)	
poly(n, 2)2			−12.600**	
			(0.442)	
log(n)				22.438**
				(3.282)
Constant	39.494	54.002**	81.250***	−109.259*
	(8.505)	(6.485)	(0.221)	(28.128)
Observations	4	4	4	4
R^2	0.988	0.944	1.000	0.959
Adjusted R^2	0.965	0.915	1.000	0.938
Residual Std. Error	5.735 (df = 1)	8.915 (df = 2)	0.442 (df = 1)	7.606 (df = 2)
F Statistic	42.344 (df = 2; 1)	33.467** (df = 1; 2)	7,198.410*** (df = 2; 1)	46.726** (df = 1; 2)
Note:				*p<0.1; **p<0.05; ***p<0.01