

Cyber Crime Trend India

Jaswinderpal Singh

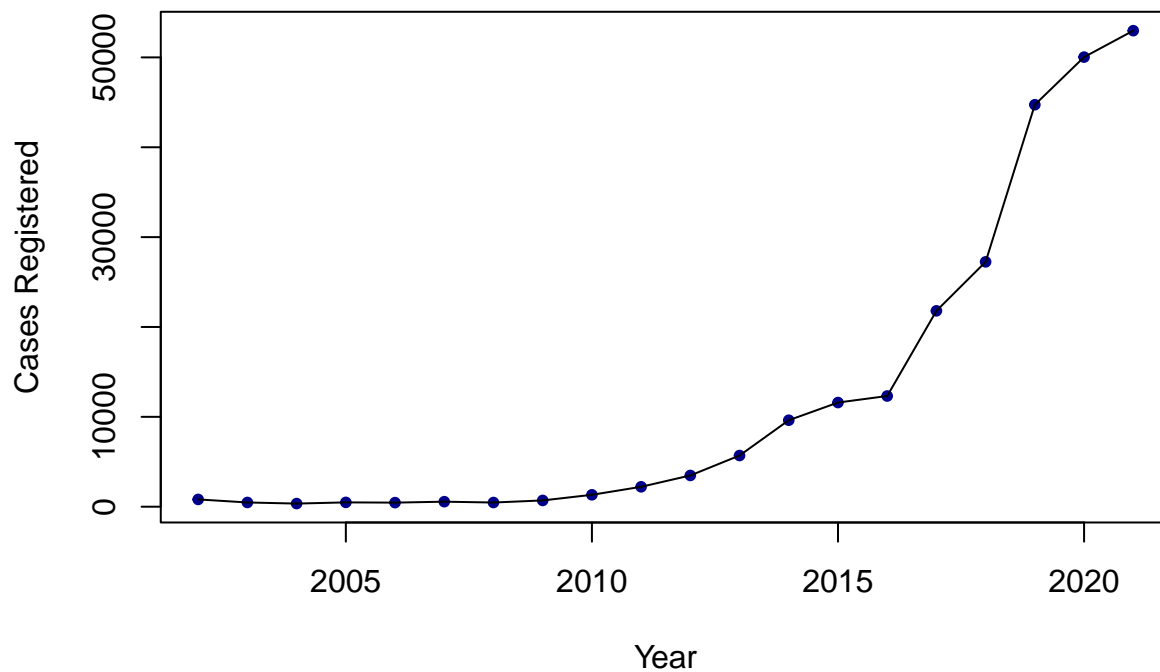
Reading trend data

```
rm(list = ls())

data = read.csv("trenddata.csv")
y = data$creg
x = data$year
rate = round(data$creg/data$midpopln, 4)
cbind(data, rate)
```

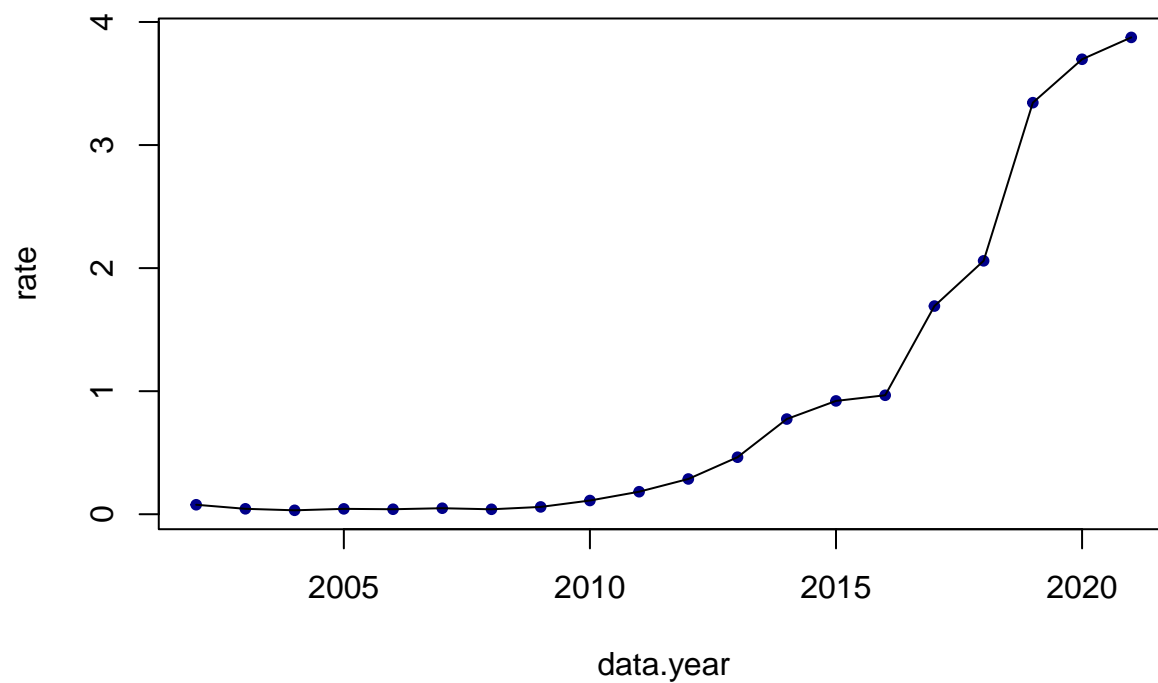
```
##   year  creg midpopln  rate
## 1  2002   808 10506.40 0.0769
## 2  2003   471 10682.14 0.0441
## 3  2004   347 10856.00 0.0320
## 4  2005   481 11027.91 0.0436
## 5  2006   453 11197.75 0.0405
## 6  2007   556 11365.53 0.0489
## 7  2008   464 11531.26 0.0402
## 8  2009   696 11694.44 0.0595
## 9  2010  1322 11857.56 0.1115
## 10 2011  2213 12101.93 0.1829
## 11 2012  3477 12133.70 0.2866
## 12 2013  5693 12287.85 0.4633
## 13 2014  9622 12440.40 0.7734
## 14 2015 11592 12591.10 0.9207
## 15 2016 12317 12739.90 0.9668
## 16 2017 21796 12885.30 1.6915
## 17 2018 27248 13233.80 2.0590
## 18 2019 44735 13376.10 3.3444
## 19 2020 50035 13533.90 3.6970
## 20 2021 52974 13671.60 3.8747
```

```
plot(data[,1:2], pch = 20, col = "blue4", xlab = "Year", ylab = "Cases Registered")
lines(data[,1:2])
```



Rate Plots

```
rda = data.frame(data$year, rate)
plot(rda, pch = 20, col = "blue4")
lines(rda)
```



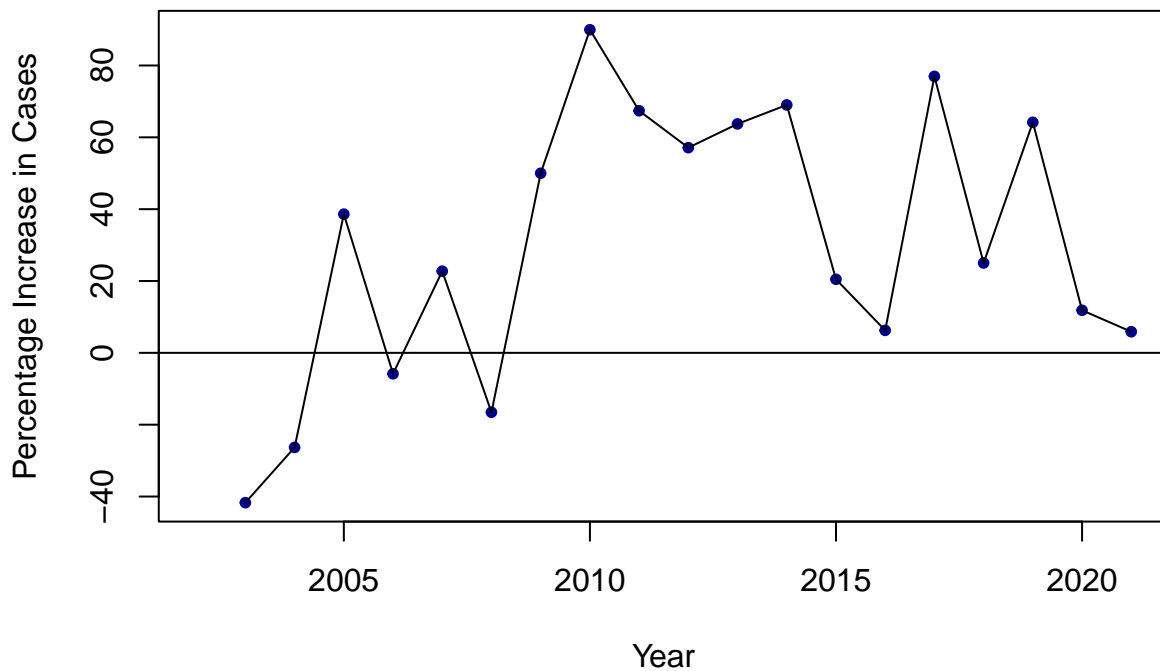
Percentage Increase

```
source("project_funcs.R")
df = cbind(data, per_inc = round(pinc(data$creg),4) )
df
```

```
##   year  creg midpopln  per_inc
```

```
## 1 2002 808 10506.40 NA
## 2 2003 471 10682.14 -41.7079
## 3 2004 347 10856.00 -26.3270
## 4 2005 481 11027.91 38.6167
## 5 2006 453 11197.75 -5.8212
## 6 2007 556 11365.53 22.7373
## 7 2008 464 11531.26 -16.5468
## 8 2009 696 11694.44 50.0000
## 9 2010 1322 11857.56 89.9425
## 10 2011 2213 12101.93 67.3979
## 11 2012 3477 12133.70 57.1170
## 12 2013 5693 12287.85 63.7331
## 13 2014 9622 12440.40 69.0146
## 14 2015 11592 12591.10 20.4739
## 15 2016 12317 12739.90 6.2543
## 16 2017 21796 12885.30 76.9587
## 17 2018 27248 13233.80 25.0138
## 18 2019 44735 13376.10 64.1772
## 19 2020 50035 13533.90 11.8475
## 20 2021 52974 13671.60 5.8739
```

```
pdat = data.frame(year = df$year, per_inc = df$per_inc)
plot(pdat, pch = 20, col = "blue4", xlab = "Year", ylab = "Percentage Increase in Cases")
lines(pdat)
abline(h = 0)
```



Analysis

```
model1 = lm(y~x)
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ x)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11298  -8925  -2259    7388   16859
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5016473     768345  -6.529 3.89e-06 ***
## x              2500         382    6.545 3.76e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9850 on 18 degrees of freedom
## Multiple R-squared:  0.7041, Adjusted R-squared:  0.6877
## F-statistic: 42.84 on 1 and 18 DF,  p-value: 3.764e-06
model2 = lm(y~ poly(x,2,raw = TRUE) )
summary(model2)
```

```
##
## Call:
## lm(formula = y ~ poly(x, 2, raw = TRUE))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7484.2 -2404.7    60.1  2556.5  6871.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.182e+09  1.136e+08   10.40 8.64e-09 ***
## poly(x, 2, raw = TRUE)1 -1.178e+06  1.130e+05  -10.43 8.38e-09 ***
## poly(x, 2, raw = TRUE)2  2.934e+02  2.808e+01   10.45 8.12e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3721 on 17 degrees of freedom
## Multiple R-squared:  0.9601, Adjusted R-squared:  0.9554
## F-statistic: 204.7 on 2 and 17 DF,  p-value: 1.275e-12
model3 = lm(log(y)~ x )
summary(model3)
```

```
##
## Call:
## lm(formula = log(y) ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.93246 -0.28671 -0.01779  0.21533  1.40482
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -589.50355    40.72045  -14.48 2.33e-11 ***
## x              0.29710     0.02024   14.68 1.86e-11 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.522 on 18 degrees of freedom
## Multiple R-squared:  0.9229, Adjusted R-squared:  0.9186
## F-statistic: 215.4 on 1 and 18 DF,  p-value: 1.856e-11
```

Predictions

```
x_axis = 2002:2021
```

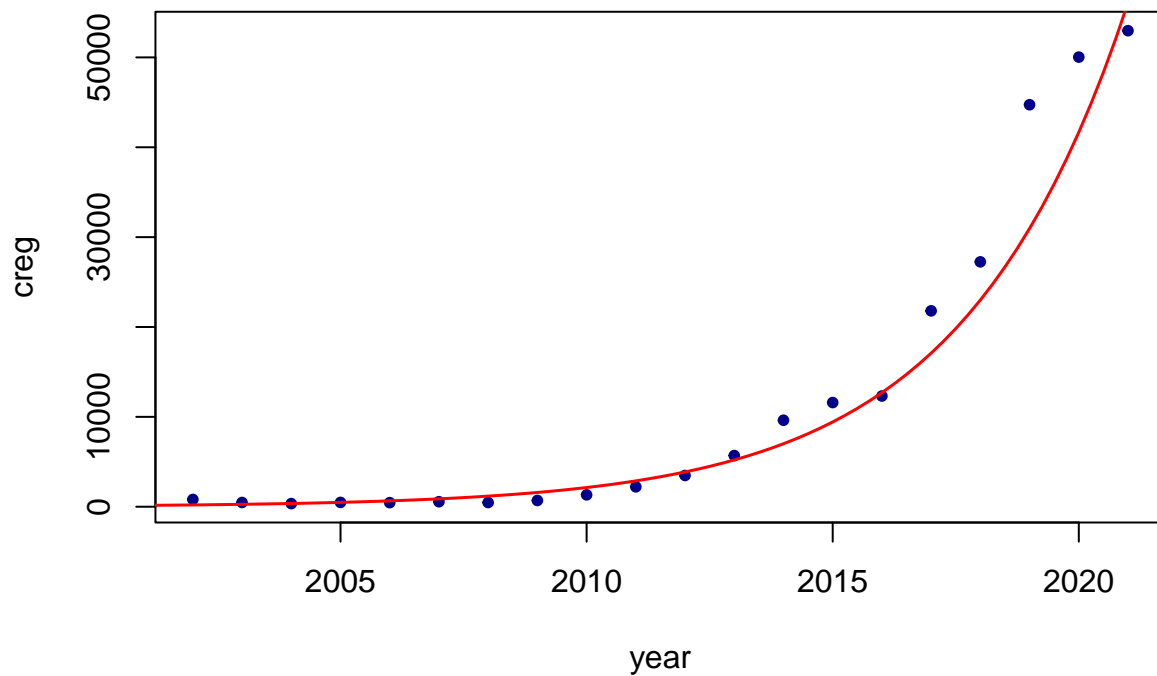
```
data.frame(year = data$year, actual = data$creg, linear = predict(model1, data.frame(x = x_axis)), quad
```

##	year	actual	linear	quad	loge
## 1	2002	808	-11385.414	5337.5143	198.2928
## 2	2003	471	-8885.371	2556.6331	266.8920
## 3	2004	347	-6385.327	362.5213	359.2231
## 4	2005	481	-3885.283	-1244.8211	483.4961
## 5	2006	453	-1385.240	-2265.3940	650.7612
## 6	2007	556	1114.804	-2699.1975	875.8916
## 7	2008	464	3614.847	-2546.2316	1178.9057
## 8	2009	696	6114.891	-1806.4962	1586.7473
## 9	2010	1322	8614.935	-479.9915	2135.6814
## 10	2011	2213	11114.978	1433.2827	2874.5188
## 11	2012	3477	13615.022	3933.3263	3868.9565
## 12	2013	5693	16115.065	7020.1393	5207.4192
## 13	2014	9622	18615.109	10693.7218	7008.9221
## 14	2015	11592	21115.153	14954.0737	9433.6537
## 15	2016	12317	23615.196	19801.1950	12697.2194
## 16	2017	21796	26115.240	25235.0857	17089.8132
## 17	2018	27248	28615.283	31255.7459	23002.0218
## 18	2019	44735	31115.327	37863.1754	30959.5546
## 19	2020	50035	33615.371	45057.3744	41669.9902
## 20	2021	52974	36115.414	52838.3429	56085.6933

Fits

```
plot(data[,1:2], pch = 20, col = "blue4")
```

```
curve( exp(predict(model3, data.frame(x))), from = 2000, to = 2025, add = TRUE, lwd = 1.5, col = "red")
```



```
cregis = exp(predict(model3, data.frame(x = 2022:2030)) )
cregis = c(data$cregis[x == 2021] , cregis)
data.frame(year = 2021:2030, cregis )
```

```
##   year   cregis
## 2021 52974.0
## 1 2022 75488.5
## 2 2023 101603.7
## 3 2024 136753.4
## 4 2025 184063.2
## 5 2026 247739.7
## 6 2027 333445.0
## 7 2028 448800.1
## 8 2029 604062.2
## 9 2030 813037.1
```