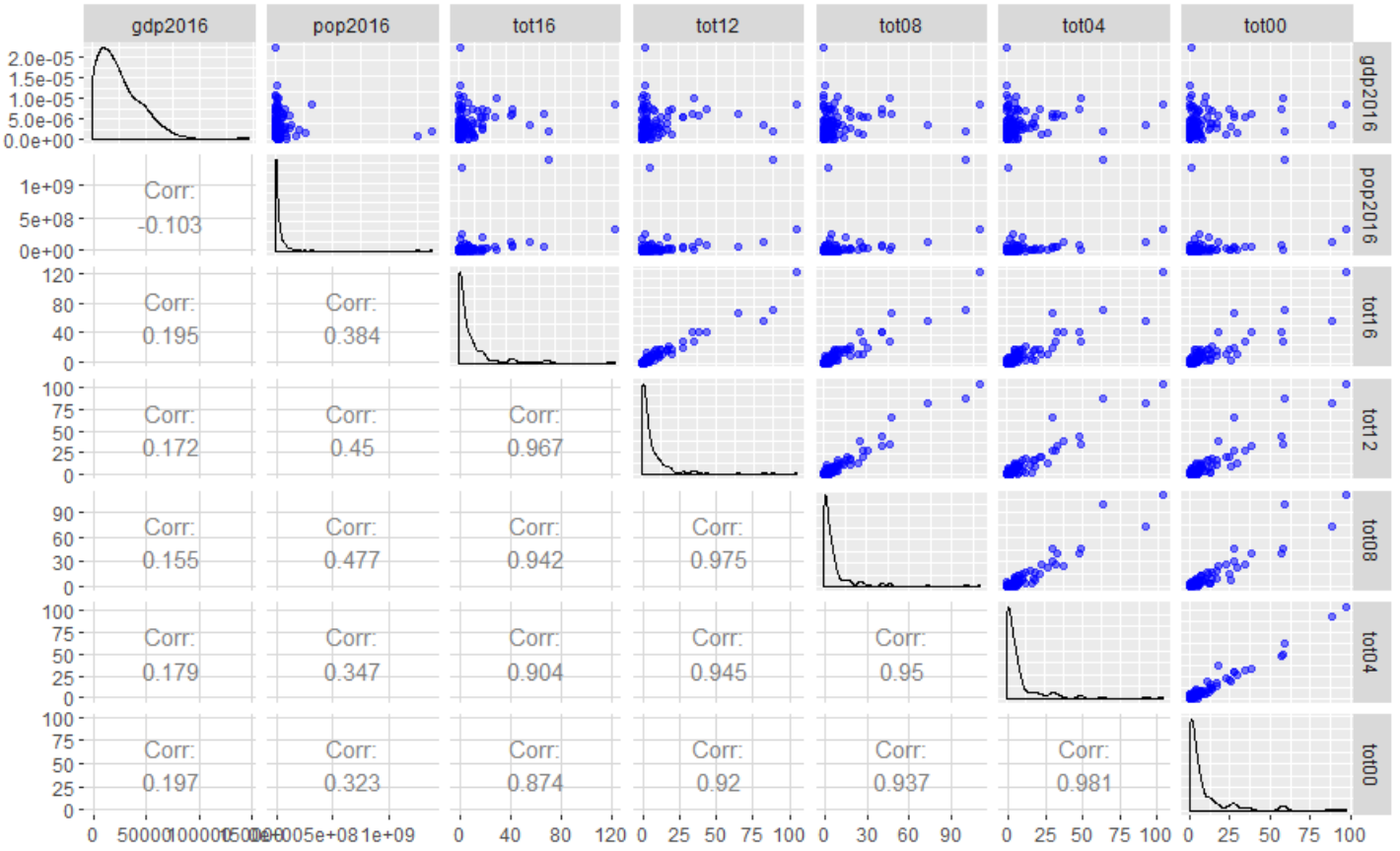
Olympics Medal Predictions
Rio 2016

Exploratory Data Analysis



Modelling

- Training data set (2012 data)
- Testing data set (2016 data)
- Response distribution
    - Normal distribution (linear model)
    - Poisson distribution .
    - Quasi-Poisson distribution
    - Negative Binomial distribution
    - Zero-inflated Poisson or negative binomial distribution

Normal Linear Model

- Most important predictor is total medal for the previous olympics.
- In sample model performance

  MAE MFE RMSE

  2.37 0.22 3.85
- Out of sample performance

  MAE MFE RMSE

  2.39 0.21 4.37
- Although the performance is relatively good but assumptions are dubious

GLM (poisson distribution)

- Model assumptions.

  Based on deviance and residual df, Lack Goodness-of-fit.

  - Overdispersion.
  - Around 20% of the response variable are zeros.
- Solutions: Quasi-Poisson ,negative binomial and zero-inflated Poisson or negative binomial.

Models Results

st.error	Pois	Qpois	Neg	Zin
(Intercept)	1.855	3.346	3.026	2.994
log(GDPpercapita)	0.062	0.11	0.102	0.101
log(pop)	0.085	0.156	0.143	0.141
host1	0.132	0.258	0.718	0.193
comm1	0.084	0.192	0.428	0.615
muslim1	0.236	0.468	0.325	0.436
muslim2	0.215	0.424	0.206	0.302
totprev1	0.002	0.003	0.005	0.006
	0	0	0	0

Models Reults

	Pois	Qpois	Neg	Zin
logLik	-345	NA	-266	-260
Df	8	8	9	14

	Pois	Qpois	Neg	Zin
AIC	706.600	NA	550.749	548.920

	Zeros Count		Zeros Count
Train	21	Test	25
Pois	8	Pois	9
NB	18	NB	17
ZINB	22	ZINB	22

In Sample Performance

poiss.RMSE	Qpoiss.RMSE	nb.RMSE	zin.RMSE
7.795	6.448	78.766	45.978
poiss.MAE	Qpoiss.MAE	nb.MAE	zin.MAE
4.453	4.465	15.132	9.925
poiss.MFE	Qpoiss.MFE	nb.MFE	zin.MFE
0.037	0.037	-10.485	-5.642

Out Sample Performance

poiss.MAE	Qpoiss.MAE	nb.MAE	zin.MAE
4.792	4.704	11.700	8.416
poiss.MFE	Qpoiss.MFE	nb.MFE	zin.MFE
0.280	0.186	-7.473	-4.115
poiss.RMSE	Qpoiss.RMSE	nb.RMSE	zin.RMSE
7.591	7.406	53.244	32.930

Quassi-Poisson and Negative Binomial