STAT 526 Group Project: Association Study of ETFs & their volatility indices.

- **Group members:** Li, Jianing, Raman, Krishnan, Wang, Anqi, Xu, Qianli:
- **Problem Description:** We investigate the relationship between exchange traded funds & their corresponding volatility indicators (vix, vxd and vxn) in the US stock market.
- **Summary of the data:** We obtained daily market dataset pertaining to 3 ETFs for 2001-2018 from Wharton/CRSP. A tiny snapshot of the data (original data 5000 rows x 13 cols):

Date	vix	vxn	vxd	spy.PERMNO	spy.PRC	spy.OPENPRC	dia.PERMNO	dia.PRC	dia.OPENPRC	qqq.PERMNO	qqq.PRC	qqq.OPENPRC
2/2/2001	21.95	54.89	19.94	84398	134.8	137.39999	85765	108.63	109.8	86755	61.55	64.94
2/5/2001	22.19	55.85	19.98	84398	135.78999	134.8	85765	109.8	108.87	86755	61.5	61.19
2/6/2001	21.98	53.68	19.57	84398	135.39	135.3	85765	109.58	109.3	86755	61.6	61.49
2/7/2001	21.67	54.41	19.2	84398	134.69	134.72	85765	109.62	109.49	86755	60.6	60.51

  - $X \in \{SPY, DIA, QQQ\}$: "Spiders" (SPY), "Diamonds" (DIA) and "Qubes" (QQQ) are the three most popular ETFs (exchange trade funds), which are the weighted sum of stocks using indices of S&P, Dow Jones and NASDSQ respectively.
  - $X.PRC_t$ = the close price of X on date t.
  - Monthly Return(X) = (X(t+30) - X(t))/X(t)
  - $v \in \{vix, vxd, vxn\}$: Imply the volatility of SPY, DIA and QQQ respectively, which are good estimators of short-term standard deviation of X (i.e., $\hat{\sigma}_X$).
  - v_t = the close "prices" of v on date t.
    - Monthly Return(v) = (v(t+30) - v(t))/v(t)
- **Methods:** For each of the 3 ETF-volatility pairs, we convert continuous price data into ordinal counts by counting the number of days when ETF & vol Returns fall into distinct Monthly Return buckets. The cell-counts yield a contingency table as shown below:

	QQQ < -10%	*QQQ < -5%*	*QQQ < 0%*	*QQQ < 5%*	*QQQ < 10%*
Vxn < -10%					
Vxn < -5%					
Vxn < 0%					
Vxn < 5%					
Vxn < 10%					

Based on the contingency table, we can run Pearson to test statistical independence, compute odds ratios along compressed 2x2 tables, build proportional odds logistic regression models, multinomial & cumulative logit models, compare among models using residual deviances, test goodness of fit from predicted cell counts, build multiple loglinear models using GLMs etc.