# PCA On Stock Price

Group2: Yaxin Li, Blake Hillier, Joe Puhalla

January 2020

This project analyzes the relationships multiple stock prices have on each other. We pulled 100 stocks from Bloomberg using daily data from January 2020 back until the 1970's-1980's depending on the stock. We then randomly selected three stocks from the list below, removing rows until each stock had the same amount of data, and then applied the PCA method to the data to find the singular values. This was repeated until we had 33 combinations with no stock used twice.

The 100 stocks are:

abt	adp	aep	aig	axp	ba	bac	bax	bdx	bk
bmy	bp	cat	cl	clx	cmi	cop	cpb	csx	cvs
cvx	d	de	dhr	duk	eix	emr	es	etn	etr
exc	fdx	fitb	gd	ge1	gis	glw	gpc	gww	hal
hban	hpq	hsy	hum	ibm1	iff	intc	ip	jnj	jpm
k	klac	kmb	ko	kr	l	leg	lhx	lly	lnc
mkc	mlhr	mmm	mo	mrk	msi	mtb	nem	nsc	ntrs
omc	pcar	peg	pep	pfe	pg	pnw	ppg	ppl	rtn
shw	slb	so	spgi	tgt	tmo	tsn	txn	unp	usb
utx	vfc	wba	wfc	whr	wmb	wy	xel	xom	xrx

Singular values for each combination
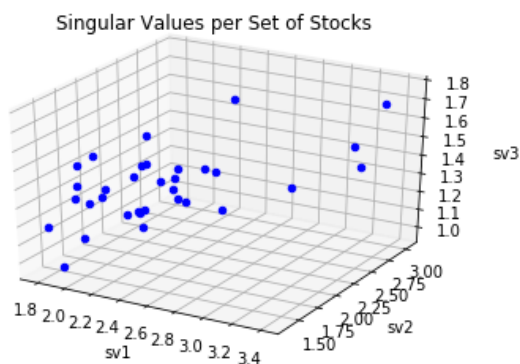
Once we obtained all the singular values, we plotted them to determine the number of clusters for KMeans.

Each color represents a different cluster
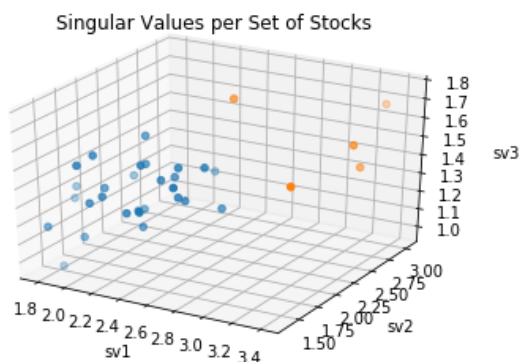
We then applied KMeans to the data with $K = 2$.