Impute evaluation vignette

Package install and data import

```
library(magrittr)
library(knitr)
library(ggplot2)
library(reshape2)
source('Imputation evaluations.R')
data_test <- read.csv('OB_data/Ob_met_nona.csv', row.names = 1)
group <- rownames(data_test) %>% gsub('()-.*', '\\1', .) %>% as.factor()
```

	UDCA	CDCA	CA	GCDCA	GHCA	GCA	TCDCA	TCA	Primary	Secondary
DM-1	70.791	514.177	206.219	411.484	4.721	21.626	7.300	0.648	1161.454	151.6205
DM-10	50.623	128.610	41.453	227.880	0.540	47.970	4.781	2.123	452.817	312.4948
DM-11	42.720	125.704	46.343	160.441	2.990	9.155	9.994	0.524	352.161	304.2185
DM-12	15.682	67.553	21.916	71.591	0.728	23.039	5.737	2.506	192.342	217.5495
DM-13	65.920	72.615	18.242	123.881	1.381	18.932	3.326	0.708	237.704	156.0188
DM-14	161.750	379.706	26.088	274.699	0.490	43.720	15.053	4.504	743.770	215.1234

```
##   [1] DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM
##  [24] DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM
##  [47] DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM DM
##  [70] DM N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N
##  [93] N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N
## [116] N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N
## [139] N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N
## [162] N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N
## [185] N  N  N  N  N  N  N  N  N  N  N  N  N  N
## Levels: DM N
```

MAR

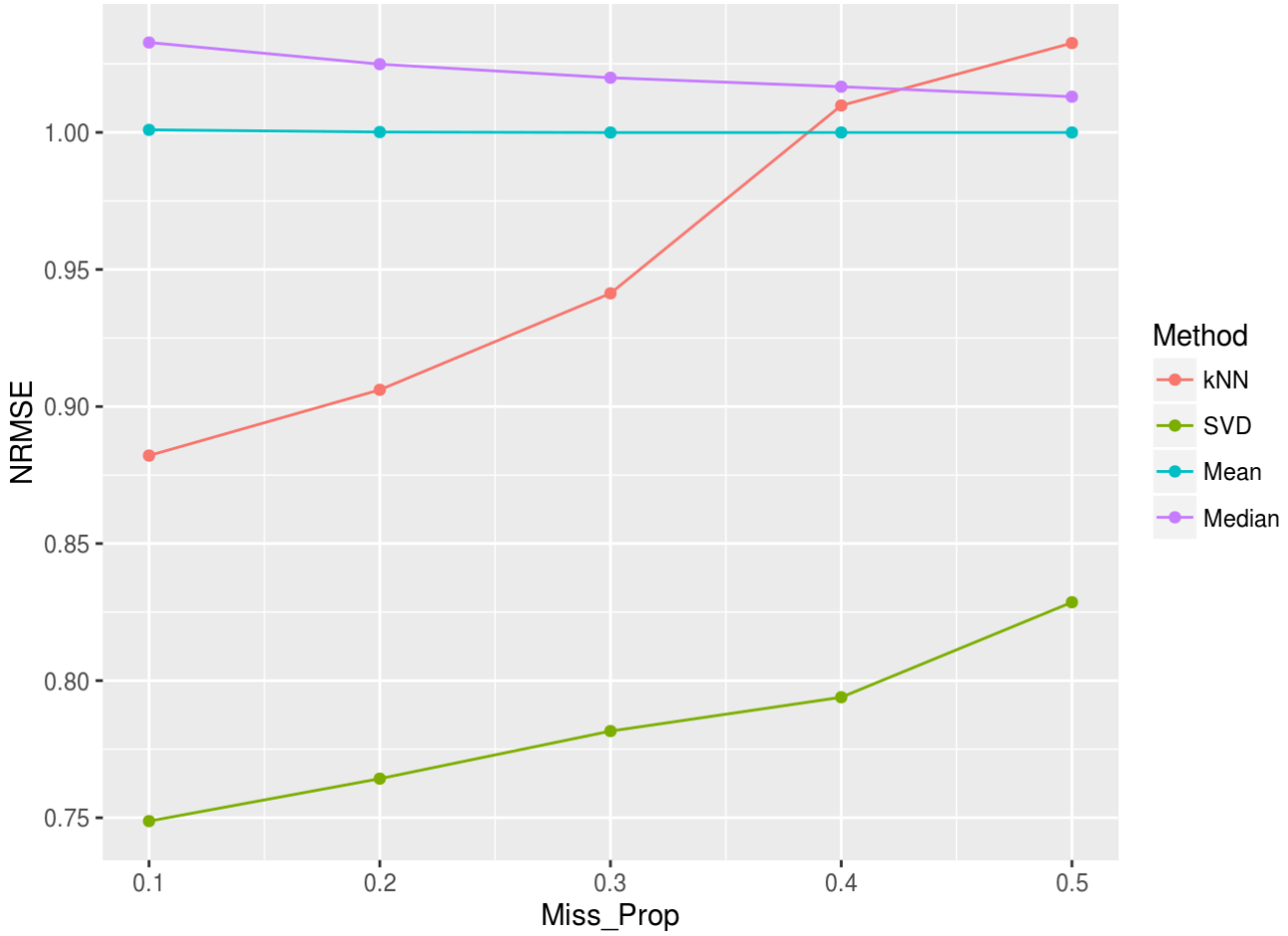
MAR generation and imputation

```
MAR_list <- MAR_gen_imp(data_c = data_test, prop = seq(.1, .5, .1), impute_list = c('kNN_wrappe
r', 'SVD_wrapper', 'Mean_wrapper', 'Median_wrapper'), cores = 5)
```

MAR NRMSE evaluation and plot

```
MAR_NRMSE_list <- NRMSE_cal_plot(MAR_list, plot = T, x = 'Miss_Prop')
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```



kNN	SVD	Mean	Median	Miss_Prop	Miss_Num
0.8821190	0.7487686	1.0009154	1.032784	0.1	130
0.9061074	0.7642424	1.0001274	1.024876	0.2	130
0.9412829	0.7815976	0.9999463	1.019928	0.3	130
1.0098120	0.7939394	0.9999619	1.016650	0.4	130
1.0325871	0.8285979	0.9999635	1.013016	0.5	130

```
## The above table shows the NRMSE of different imputaion methods
```

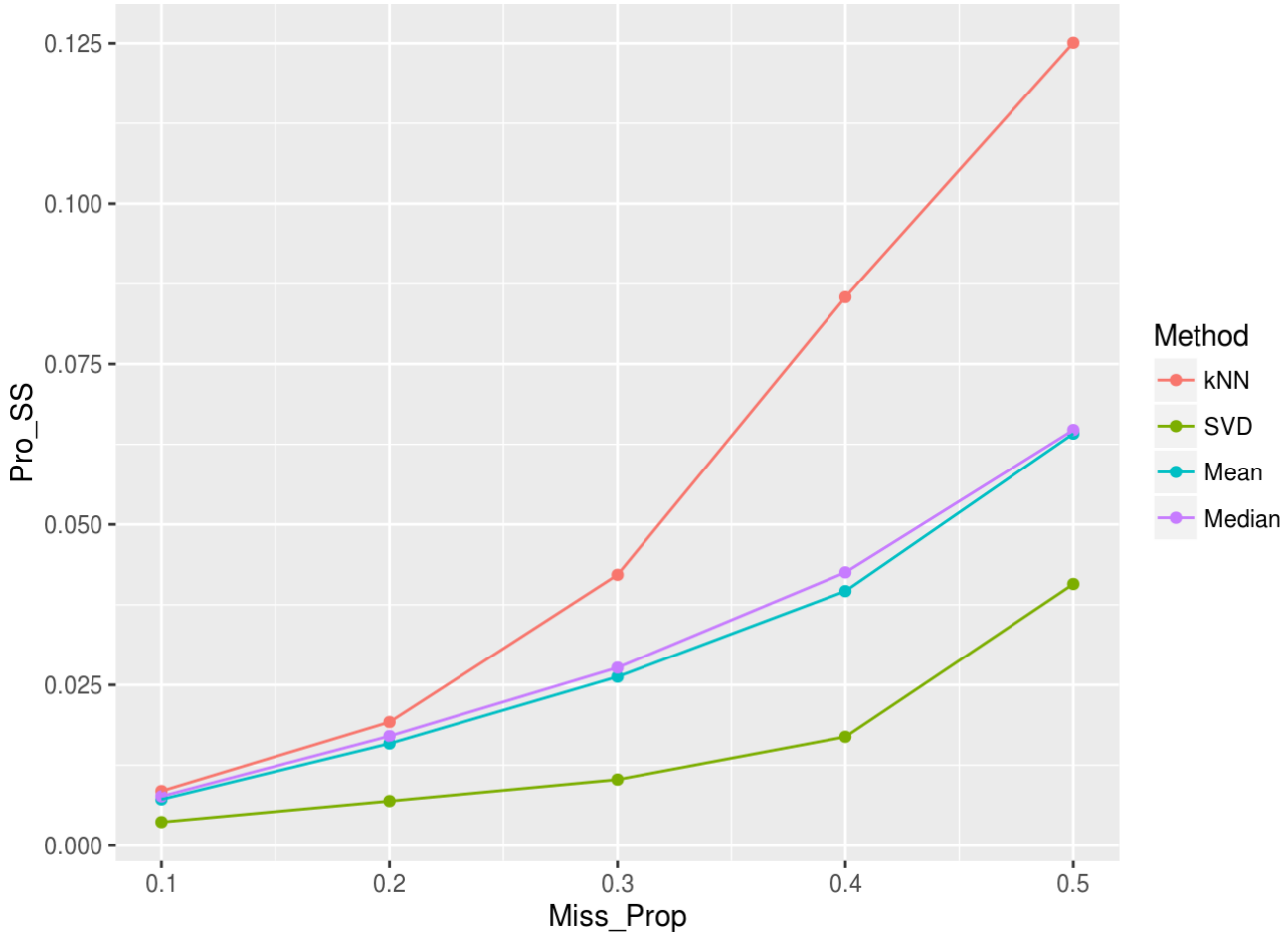
Miss_Prop	Method	NRMSE
0.1	kNN	0.8821190
0.2	kNN	0.9061074
0.3	kNN	0.9412829
0.4	kNN	1.0098120
0.5	kNN	1.0325871
0.1	SVD	0.7487686

```
## The above melted table is good for ggplot2
```

MAR PCA Procrustes analysis and plot

```
MAR_PCA_ProSS_list <- Procrustes_cal_plot(MAR_list, DR = 'PCA', nPCs = 2, x = 'Miss_Prop', plot
= T)
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```



kNN	SVD	Mean	Median	Miss_Prop	Miss_Num
0.0084779	0.0036577	0.0071913	0.0076705	0.1	130
0.0192072	0.0069267	0.0158628	0.0170239	0.2	130
0.0421603	0.0102584	0.0262739	0.0276967	0.3	130
0.0854171	0.0169050	0.0396199	0.0425408	0.4	130
0.1250814	0.0407332	0.0641691	0.0647407	0.5	130

```
## The above table shows the Procrustes Sum of Squared Error of different imputaion methods
```

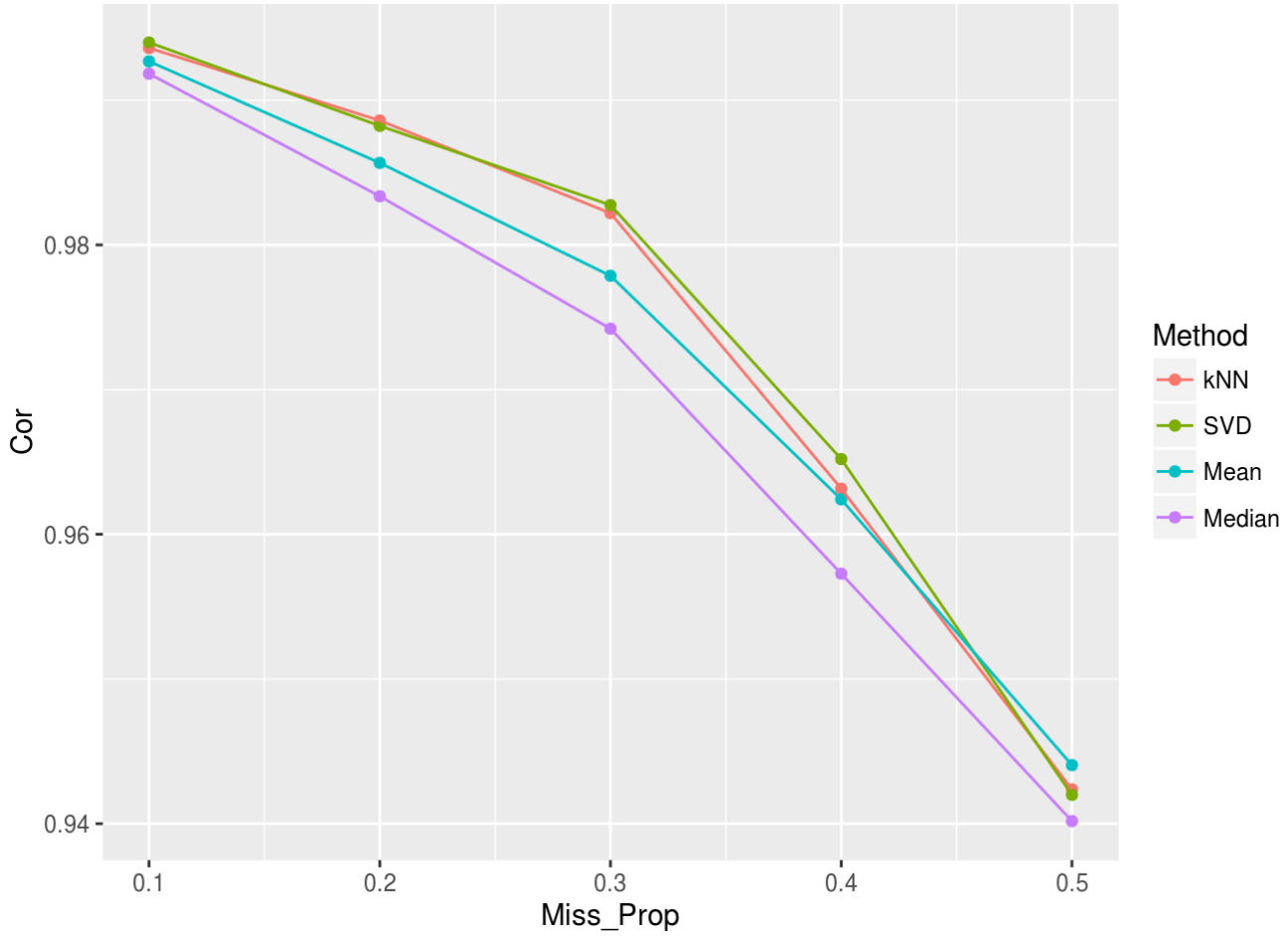
Miss_Prop	Method	Pro_SS
0.1	kNN	0.0084779
0.2	kNN	0.0192072
0.3	kNN	0.0421603
0.4	kNN	0.0854171
0.5	kNN	0.1250814
0.1	SVD	0.0036577

```
## The above melted table is good for ggplot2
```

MAR T-test results correlation

```
MAR_Ttest_cor_list <- Ttest_cor_cal_plot(MAR_list, group = group, plot = T, x = 'Miss_Prop', cor
 = 'P')
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```



kNN	SVD	Mean	Median	Miss_Prop	Miss_Num
0.9936050	0.9939870	0.9926821	0.9918314	0.1	130
0.9885944	0.9882247	0.9856688	0.9833635	0.2	130
0.9821947	0.9827547	0.9778669	0.9742084	0.3	130
0.9631565	0.9651974	0.9624143	0.9572746	0.4	130
0.9423752	0.9419902	0.9440462	0.9401788	0.5	130

```
## The above table shows the Pearson Correlation of log T-test P-values between imputed data and
 complete data
```

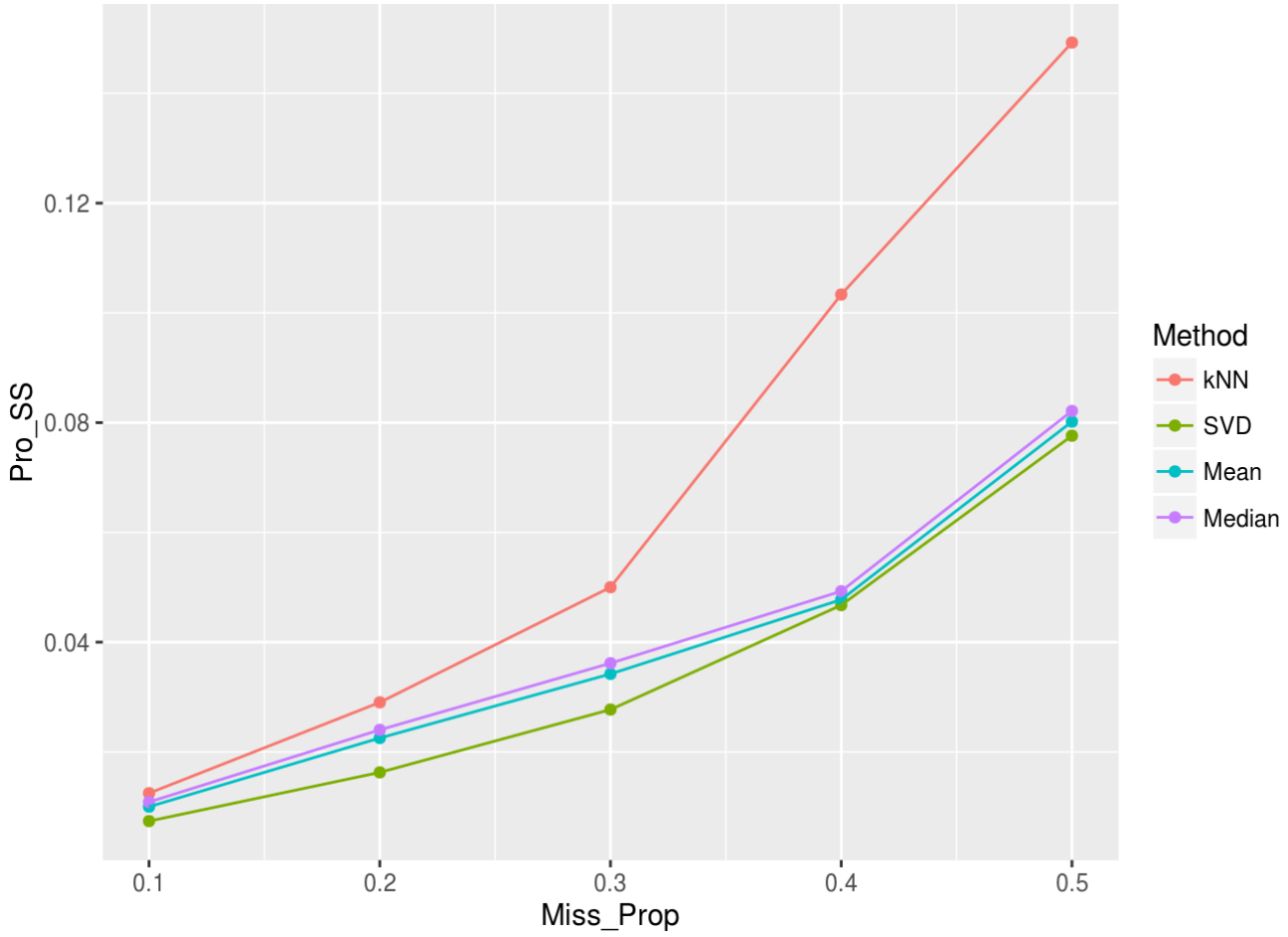
kNN	SVD	Mean	Median	Miss_Prop	Miss_Num
0.9867266	0.9869724	0.9868686	0.9863715	0.1	130
0.9814500	0.9752939	0.9753868	0.9767196	0.2	130
0.9703724	0.9629600	0.9656639	0.9656584	0.3	130
0.9452239	0.9251389	0.9425528	0.9424709	0.4	130
0.9109697	0.8998047	0.9232381	0.9222822	0.5	130

```
## The above table shows the Spearman Correlation of T-test P-values between imputed data and co
mplete data
```

MAR PLS Procrustes analysis and plot

```
MAR_PLS_ProSS_list <- Procrustes_cal_plot(MAR_list, DR = 'PLS', nPCs = 2, outcome = group, x =
'Miss_Prop', plot = T)
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```



kNN	SVD	Mean	Median	Miss_Prop	Miss_Num
0.0124798	0.0073960	0.0100185	0.0108814	0.1	130
0.0290273	0.0162707	0.0225334	0.0240209	0.2	130
0.0500077	0.0277171	0.0342118	0.0361674	0.3	130
0.1033455	0.0467446	0.0477503	0.0492796	0.4	130
0.1492585	0.0776162	0.0801839	0.0821324	0.5	130

```
## The above table shows the Procrustes Sum of Squared Error of different imputaion methods
```

Miss_Prop	Method	Pro_SS
0.1	kNN	0.0124798
0.2	kNN	0.0290273
0.3	kNN	0.0500077
0.4	kNN	0.1033455
0.5	kNN	0.1492585
0.1	SVD	0.0073960

```
## The above melted table is good for ggplot2
```

MNAR

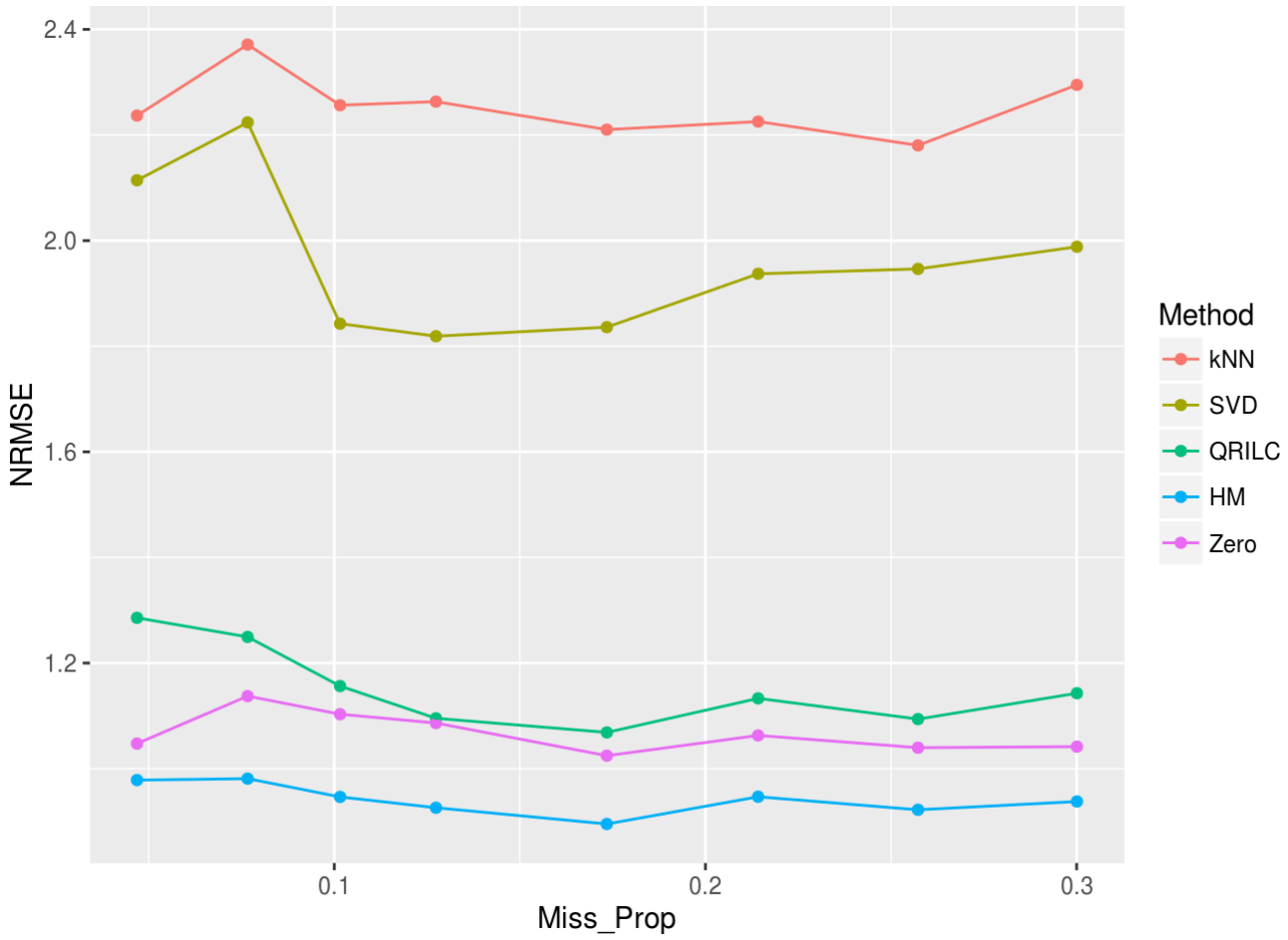
MNAR generation and imputation

```
MNAR_list <- MNAR_gen_imp(data_c = data_test, mis_var_prop = seq(.1, .8, .1), var_mis_prop =
seq(.1, .6, .01), impute_list = c('kNN_wrapper', 'SVD_wrapper', 'QRILC_wrapper', 'HM_wrapper',
'Zero_wrapper'), cores = 5)
```

MNAR NRMSE evaluation and plot

```
MNAR_NRMSE_list <- NRMSE_cal_plot(MNAR_list, plot = T, x = 'Miss_Prop')
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```



kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
2.236687	2.114244	1.285600	0.9783366	1.047413	0.0468531	13
2.371218	2.223687	1.249410	0.9810837	1.137337	0.0766511	26
2.256369	1.842634	1.156335	0.9464719	1.103015	0.1015540	39
2.263145	1.818908	1.095270	0.9259699	1.086315	0.1273893	52
2.210140	1.836025	1.068555	0.8952246	1.024492	0.1734266	65
2.225357	1.937164	1.133020	0.9467499	1.062843	0.2141803	78

```
## The above table shows the NRMSE of different imputaion methods
```

	Miss_Prop	Method	NRMSE
	0.0468531	kNN	2.236687
	0.0766511	kNN	2.371218
	0.1015540	kNN	2.256369
	0.1273893	kNN	2.263145
	0.1734266	kNN	2.210140

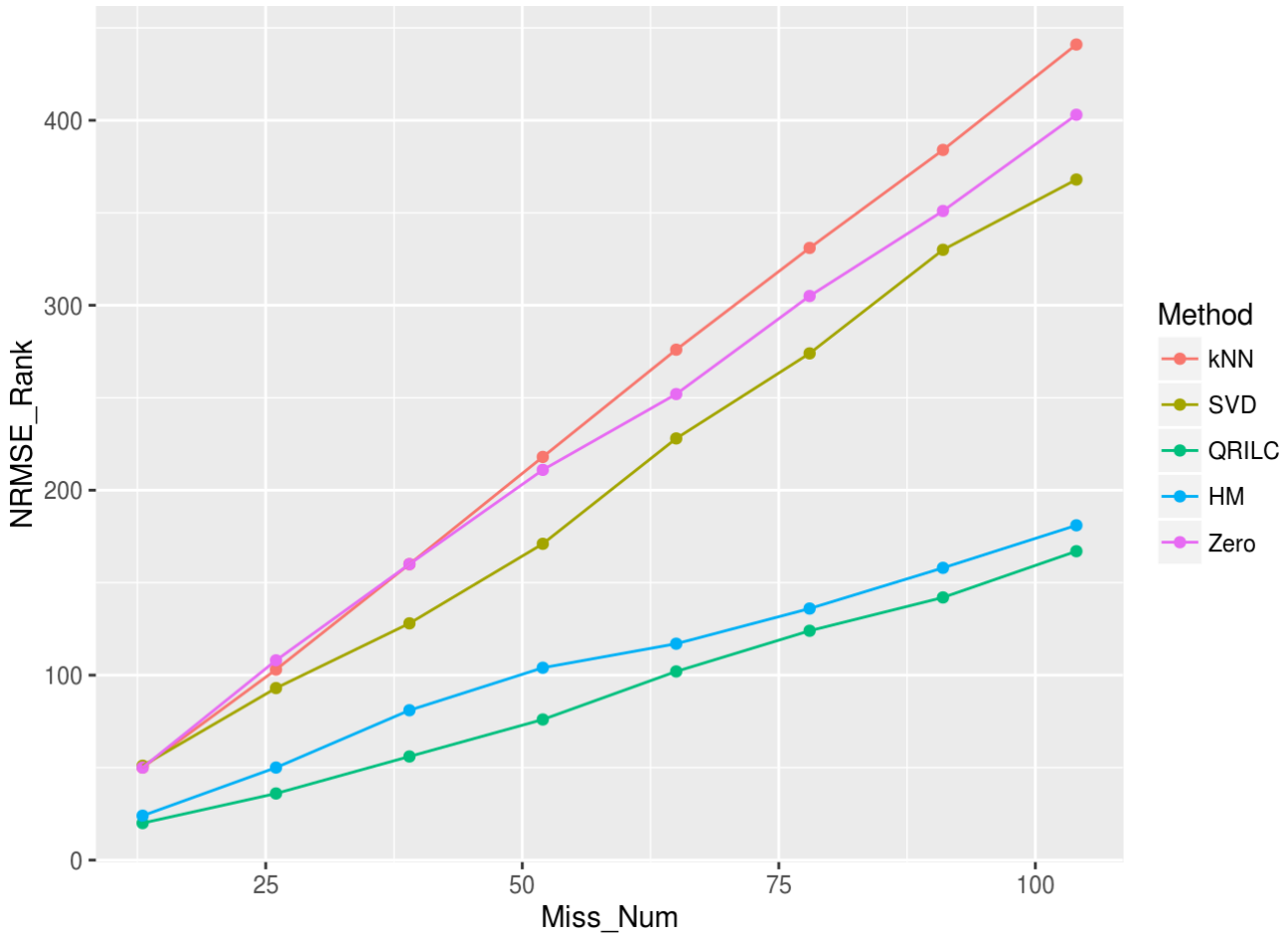
	Miss_Prop	Method	NRMSE
	0.2141803	kNN	2.225357

```
## The above melted table is good for ggplot2
```

MNAR NRMSE rank evaluation and plot

```
MNAR_NRMSE_rank_list <- NRMSE_rank_cal_plot(MNAR_list, plot = T, x = 'Miss_Num')
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```



kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
50	51	20	24	50	0.0468531	13

kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
103	93	36	50	108	0.0766511	26
160	128	56	81	160	0.1015540	39
218	171	76	104	211	0.1273893	52
276	228	102	117	252	0.1734266	65
331	274	124	136	305	0.2141803	78

```
## The above table shows the NRMSE ranks of different imputaion methods
```

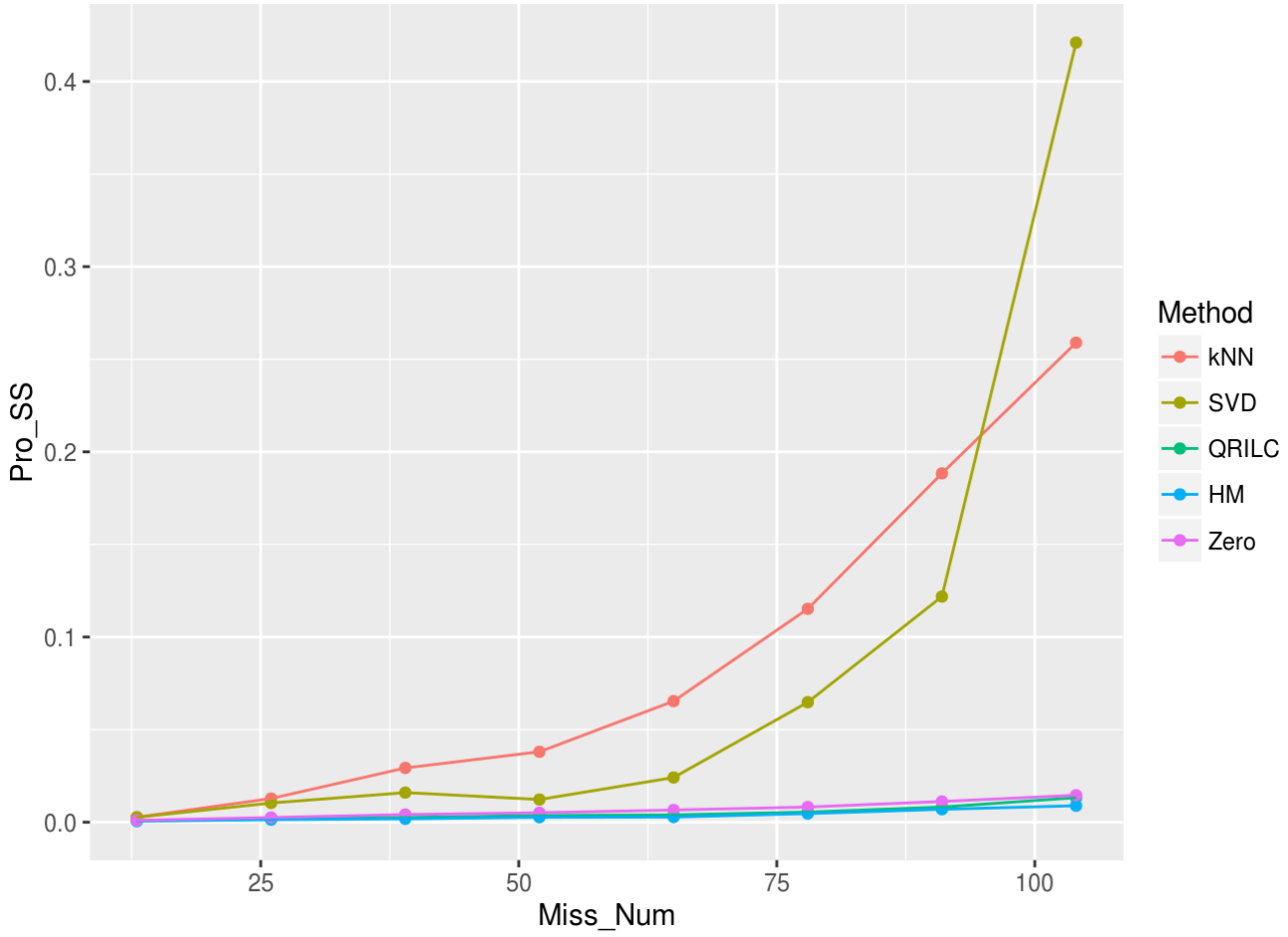
Miss_Num	Method	NRMSE_Rank
13	kNN	50
26	kNN	103
39	kNN	160
52	kNN	218
65	kNN	276
78	kNN	331

```
## The above melted table is good for ggplot2
```

MNAR PCA Procrustes analysis and plot

```
MNAR_PCA_ProSS_list <- Procrustes_cal_plot(MNAR_list, DR = 'PCA', nPCs = 2, x = 'Miss_Num', plot
 = T)
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```



kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
0.0027063	0.0026730	0.0007915	0.0005931	0.0010447	0.0468531	13
0.0127939	0.0103677	0.0018689	0.0013557	0.0025632	0.0766511	26
0.0292821	0.0159858	0.0026076	0.0018310	0.0041538	0.1015540	39
0.0380220	0.0122523	0.0036627	0.0026874	0.0051428	0.1273893	52
0.0653645	0.0241181	0.0039624	0.0028045	0.0066041	0.1734266	65
0.1152185	0.0647955	0.0054751	0.0046624	0.0082017	0.2141803	78

```
## The above table shows the Procrustes Sum of Squared Error of different imputaion methods
```

Miss_Num	Method	Pro_SS
13	kNN	0.0027063
26	kNN	0.0127939
39	kNN	0.0292821
52	kNN	0.0380220
65	kNN	0.0653645

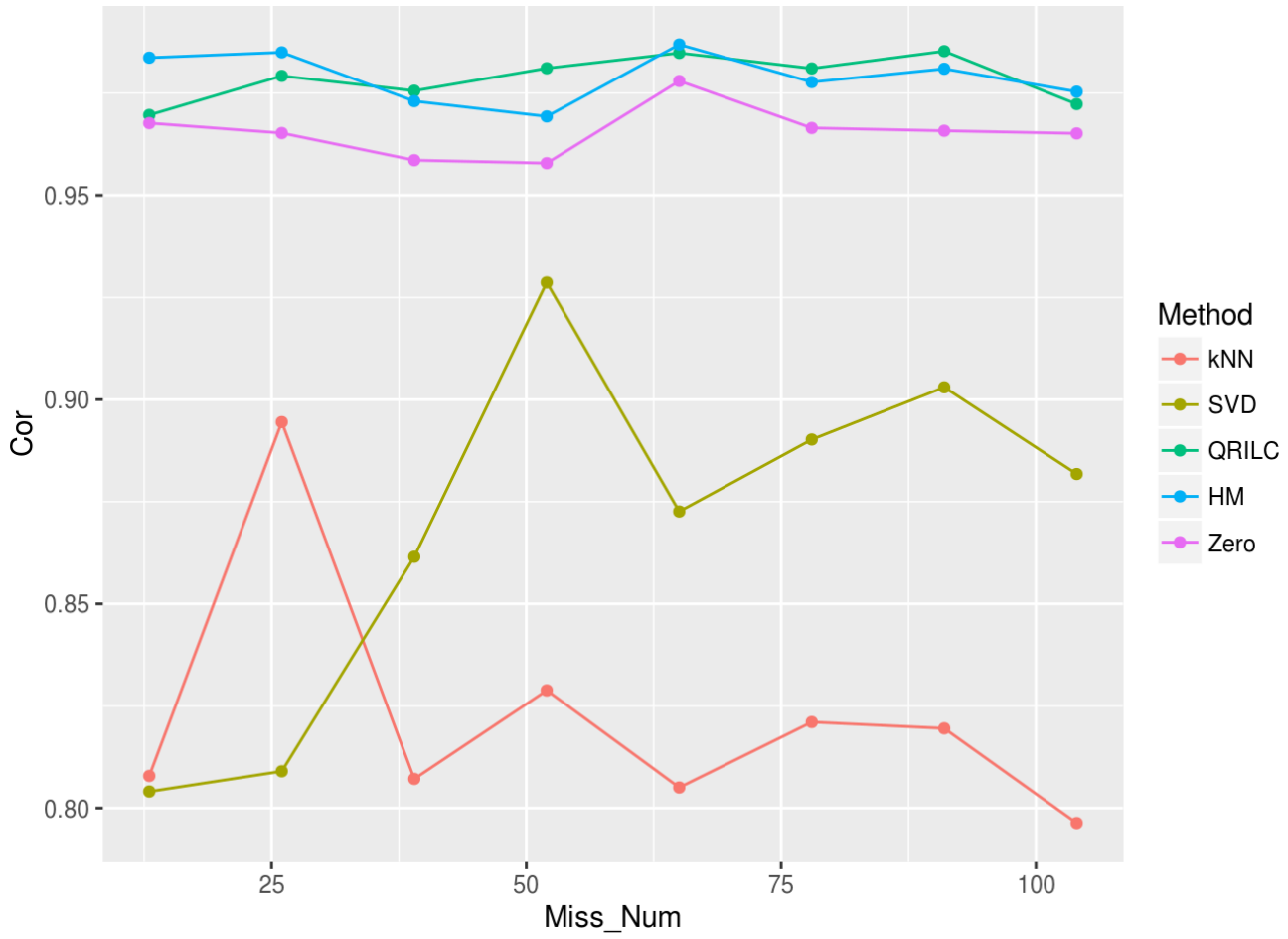
	Miss_Num	Method	Pro_SS
	78	kNN	0.1152185

```
## The above melted table is good for ggplot2
```

MNAR T-test results correlation

```
MNAR_Ttest_cor_list <- Ttest_cor_cal_plot(MNAR_list, group = group, plot = T, x = 'Miss_Num', co
r = 'P')
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```



kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
0.8078573	0.8040408	0.9696369	0.9836727	0.9676442	0.0468531	13

kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
0.8944876	0.8089978	0.9791985	0.9849790	0.9652150	0.0766511	26
0.8071283	0.8615040	0.9755600	0.9730206	0.9585651	0.1015540	39
0.8288236	0.9286601	0.9810875	0.9692799	0.9578325	0.1273893	52
0.8050209	0.8725978	0.9848267	0.9868930	0.9779477	0.1734266	65
0.8210529	0.8902189	0.9810368	0.9776853	0.9664548	0.2141803	78

```
## The above table shows the Pearson Correlation of log T-test P-values between imputed data and
 complete data
```

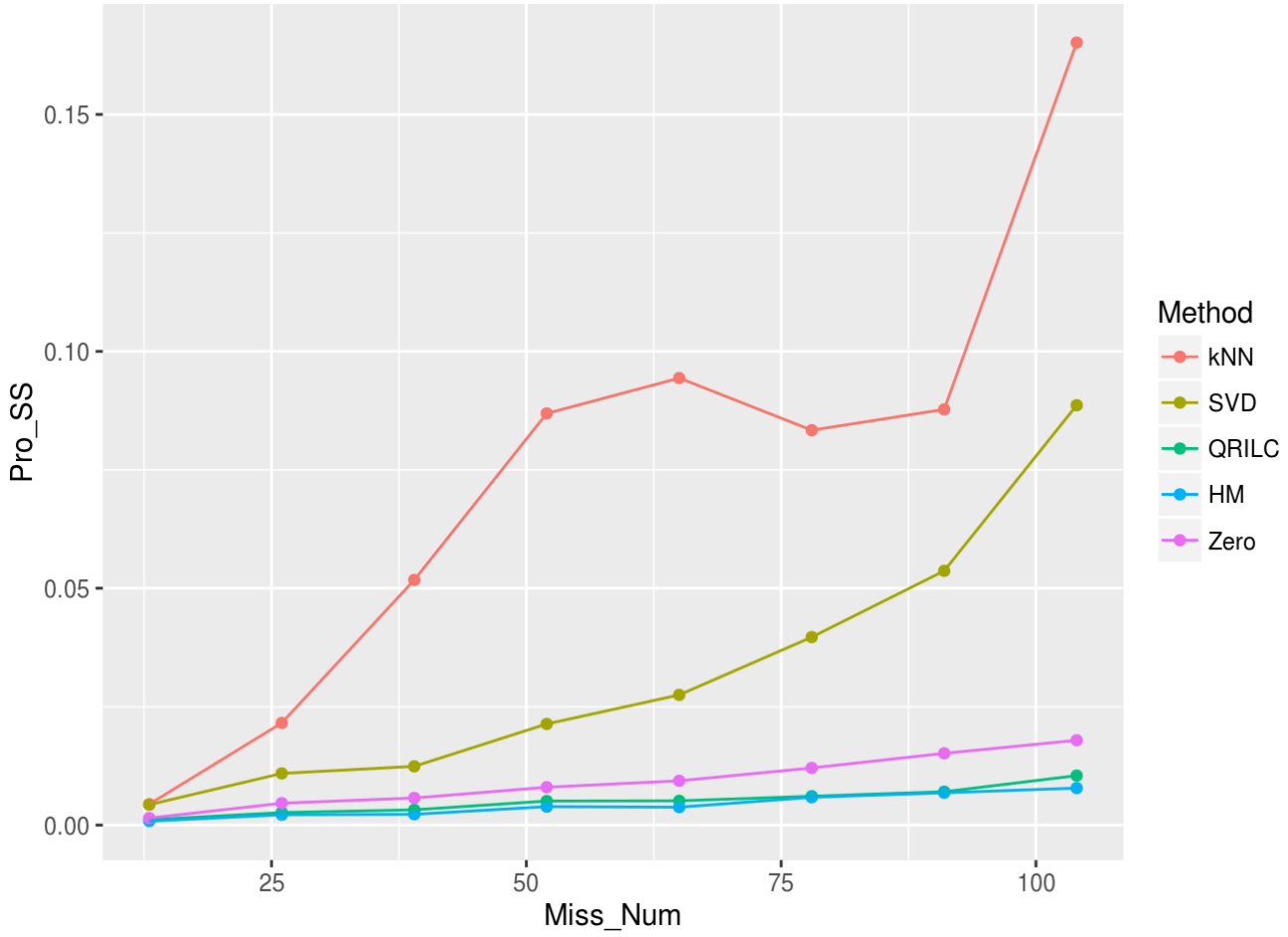
kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
0.5494505	0.5824176	0.9725275	0.9835165	0.9835165	0.0468531	13
0.8092308	0.8174359	0.9767521	0.9815385	0.9712821	0.0766511	26
0.8066802	0.8710526	0.9892713	0.9904858	0.9795547	0.1015540	39
0.8083326	0.9241014	0.9883890	0.9829250	0.9719116	0.1273893	52
0.7771853	0.8328234	0.9856206	0.9913024	0.9830857	0.1734266	65
0.7416002	0.8499222	0.9783761	0.9816639	0.9742030	0.2141803	78

```
## The above table shows the Spearman Correlation of T-test P-values between imputed data and co
mplete data
```

MNAR PLS Procrustes analysis and plot

```
MNAR_PLS_ProSS_list <- Procrustes_cal_plot(MNAR_list, DR = 'PLS', nPCs = 2, outcome = group, x =
 'Miss_Num', plot = T)
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```



kNN	SVD	QRILC	HM	Zero	Miss_Prop	Miss_Num
0.0044222	0.0043152	0.0011540	0.0008481	0.0014908	0.0468531	13
0.0215637	0.0109098	0.0026515	0.0021867	0.0046267	0.0766511	26
0.0517433	0.0124044	0.0032407	0.0022885	0.0057299	0.1015540	39
0.0869014	0.0213504	0.0050834	0.0039034	0.0079970	0.1273893	52
0.0943667	0.0274886	0.0051290	0.0037887	0.0093517	0.1734266	65
0.0833648	0.0396864	0.0061000	0.0058838	0.0120632	0.2141803	78

```
## The above table shows the Procrustes Sum of Squared Error of different imputaion methods
```

Miss_Num	Method	Pro_SS
13	kNN	0.0044222
26	kNN	0.0215637
39	kNN	0.0517433
52	kNN	0.0869014
65	kNN	0.0943667

Miss_Num	Method	Pro_SS
78	kNN	0.0833648

```
## The above melted table is good for ggplot2
```