

Important notes before you get started

1. This notebook uses SAS Software and R. R is free, SAS is not. You must have a SAS license to run SAS or use SAS onDemand for Academics (free). In addition, you need to install the SAS kernel for jupyter notebook, available here: https://github.com/sassoftware/sas_kernel
2. All file paths in this notebook will need to be revised to the destination you choose to place this notebook and all the datasets.
3. For those who don't want to install SAS but still want to have a look at the code and its output, we provide a PDF version of the notebook, including all the input and output data files from this analysis.

Dependencies: file locations and libraries

```
In [13]: %let location =F:\Manuscripts\2021_06_02_Nature_Comm_DSense\GITHUB\data;  
libname repo 'F:\Manuscripts\2021_06_02_Nature_Comm_DSense\GITHUB\data';
```

```
Out [13]: 163 ods listing close;ods html5 (id=saspy_internal) file=_tomods1 options(bitmap_mode='inlin  
e') device=svg style=HMLBLblue;  
163 ! ods graphics on / outputfmt=png;  
163 NOTE: Writing HTML5(SASPY_INTERNAL) Body file: _TOMODS1  
164  
165 %let location =F:\Manuscripts\2021_06_02_Nature_Comm_DSense\GITHUB\data;  
165 libname repo 'F:\Manuscripts\2021_06_02_Nature_Comm_DSense\GITHUB\data';  
NOTE: Libref REPO was successfully assigned as follows:  
Engine: V9  
Physical Name: F:\Manuscripts\2021_06_02_Nature_Comm_DSense\GITHUB\data  
167  
168  
169 ods html5 (id=saspy_internal) close;ods listing;  
170
```

```
In [14]: %setwd("F:\Manuscripts\2021_06_02_Nature_Comm_DSense\GITHUB\data")  
getwd()
```

```
F:\Manuscripts\2021_06_02_Nature_Comm_DSense\GITHUB\data
```

```
In [15]: library(r10)  
library(corrplot)  
library(ggplot2)  
library(ggrepel)  
library(RColorBrewer)
```

Dataset preparation

Import dataset

```
In [16]: PROC IMPORT OUT=dense datafile = "%location\dataset_bubble_plot_final.xlsx"  
DBMS = xlsx replace;  
RUN;
```

```
Out [16]: 173 ods listing close;ods html5 (id=saspy_internal) file=_tomods1 options(bitmap_mode='inlin  
e') device=svg style=HMLBLblue;  
173 ! ods graphics on / outputfmt=png;  
NOTE: Writing HTML5(SASPY_INTERNAL) Body file: _TOMODS1  
174  
175 PROC IMPORT OUT=dense datafile = "%location\dataset_bubble_plot_final.xlsx"  
176 DBMS = xlsx replace;  
177 RUN;  
  
NOTE: The import data set has 70 observations and 22 variables.  
NOTE: WORK.DSENSE data set was successfully created.  
NOTE: PROCEDURE IMPORT used (Total process time):  
real time 0.26 seconds  
cpu time 0.10 seconds  
  
178  
179  
180 ods html5 (id=saspy_internal) close;ods listing;  
181
```

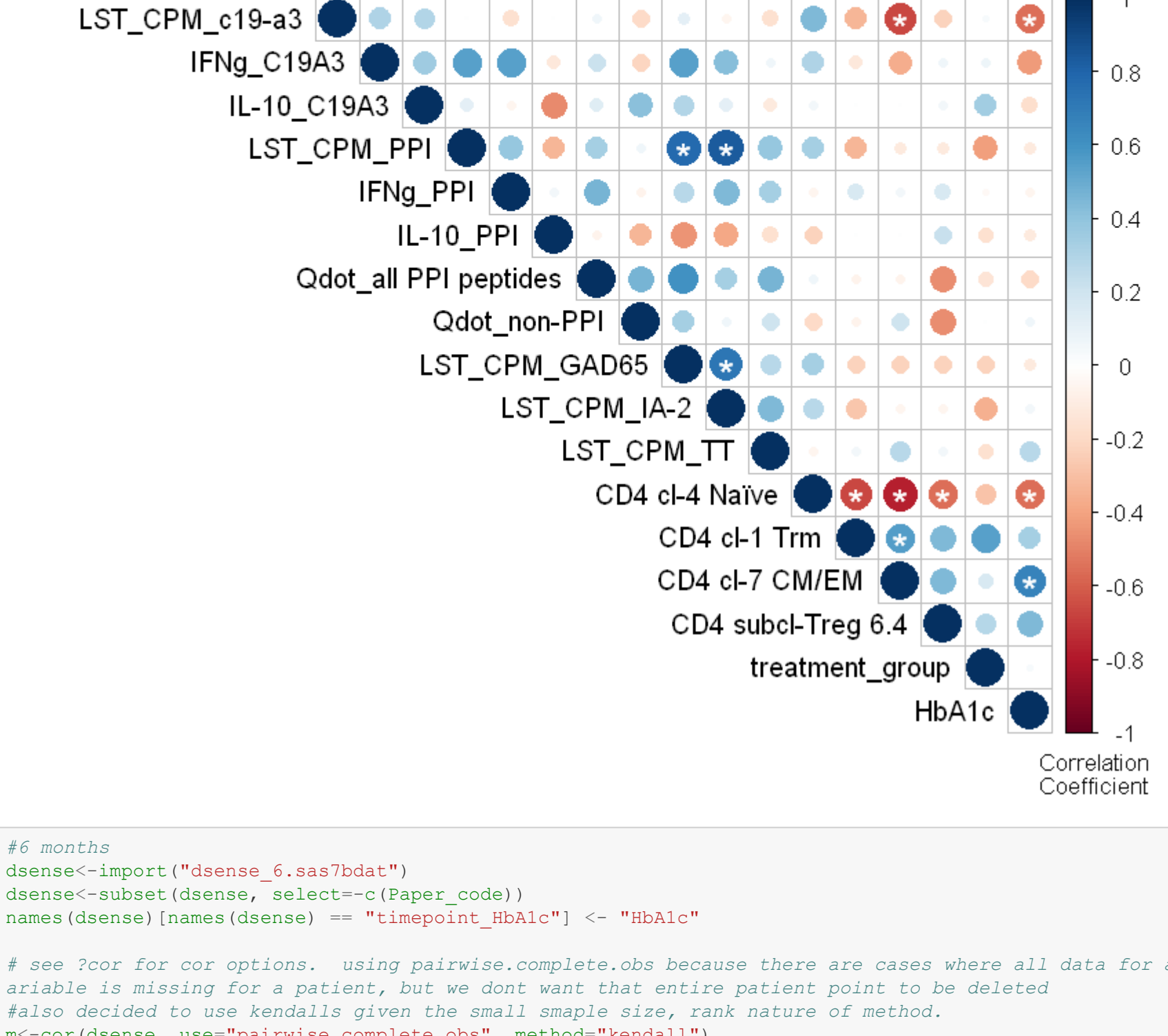
subset and export datasets for use in R

```
In [17]: DATA dsense_0 (drop=Patient pat_id Time_to_treatment_month visit);  
set dsense;  
if time_to_treatment_month=0 then delete;  
RUN;  
  
DATA repo.dsense_0;  
set dsense_0;  
RUN;  
  
DATA dsense_3 (drop=Patient pat_id Time_to_treatment_month visit);  
set dsense;  
if time_to_treatment_month=3 then delete;  
RUN;  
  
DATA repo.dsense_3;  
set dsense_3;  
RUN;  
  
DATA dsense_6 (drop=Patient pat_id Time_to_treatment_month visit);  
set dsense;  
if time_to_treatment_month=6 then delete;  
RUN;  
  
DATA repo.dsense_6;  
set dsense_6;  
RUN;  
  
DATA dsense_9 (drop=Patient pat_id Time_to_treatment_month visit);  
set dsense;  
if visit="revisit" then delete;  
RUN;  
  
DATA repo.dsense_9;  
set dsense_9;  
RUN;
```

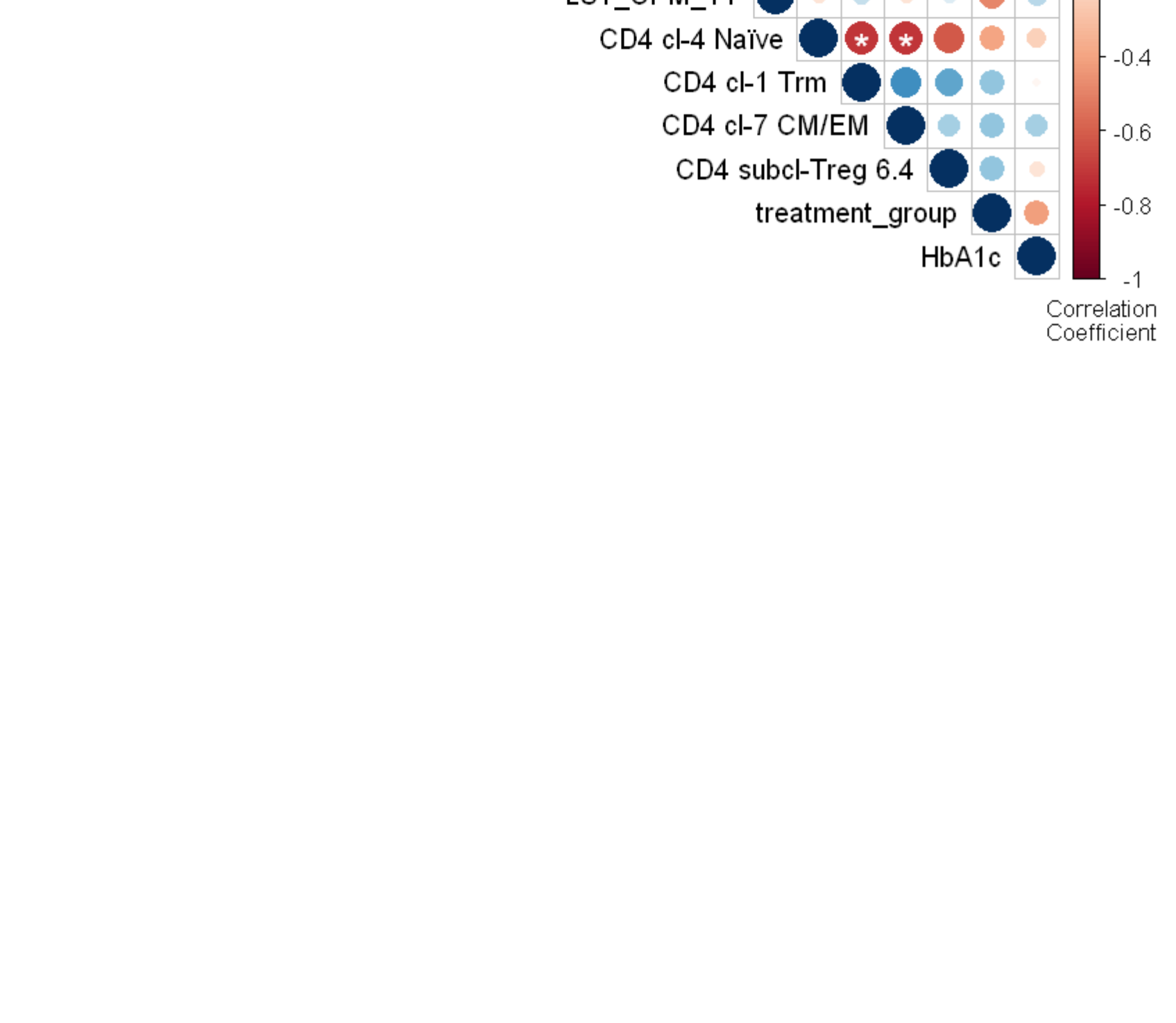
```
Out [17]: 184 ods listing close;ods html5 (id=saspy_internal) file=_tomods1 options(bitmap_mode='inlin  
e') device=svg style=HMLBLblue;  
184 ! ods graphics on / outputfmt=png;  
NOTE: Writing HTML5(SASPY_INTERNAL) Body file: _TOMODS1  
185  
186 DATA dsense_0 (drop=Patient pat_id Time_to_treatment_month visit);  
187 set dsense;  
188 if time_to_treatment_month=0 then delete;  
189 RUN;  
  
NOTE: There were 70 observations read from the data set WORK.DSENSE.  
NOTE: The data set WORK.DSENSE_0 has 9 observations and 18 variables.  
NOTE: DATA statement used (Total process time):  
real time 0.02 seconds  
cpu time 0.03 seconds  
  
190  
191 DATA repo.dsense_0;  
192 set dsense_0;  
193 RUN;  
  
NOTE: There were 9 observations read from the data set WORK.DSENSE_0.  
NOTE: The data set REPO.DSENSE_0 has 9 observations and 18 variables.  
NOTE: DATA statement used (Total process time):  
real time 1.42 seconds  
cpu time 0.01 seconds  
  
194  
195  
196 DATA dsense_3 (drop=Patient pat_id Time_to_treatment_month visit);  
197 set dsense;  
198 if time_to_treatment_month=3 then delete;  
199 RUN;  
  
NOTE: There were 70 observations read from the data set WORK.DSENSE.  
NOTE: The data set WORK.DSENSE_3 has 9 observations and 18 variables.  
NOTE: DATA statement used (Total process time):  
real time 0.04 seconds  
cpu time 0.03 seconds  
  
200  
201 DATA repo.dsense_3;  
202 set dsense_3;  
203 RUN;  
  
NOTE: There were 9 observations read from the data set WORK.DSENSE_3.  
NOTE: The data set REPO.DSENSE_3 has 9 observations and 18 variables.  
NOTE: DATA statement used (Total process time):  
real time 1.39 seconds  
cpu time 0.06 seconds  
  
204  
205 DATA dsense_6 (drop=Patient pat_id Time_to_treatment_month visit);  
206 set dsense;  
207 if time_to_treatment_month=6 then delete;  
208 RUN;  
  
NOTE: There were 70 observations read from the data set WORK.DSENSE.  
NOTE: The data set WORK.DSENSE_6 has 9 observations and 18 variables.  
NOTE: DATA statement used (Total process time):  
real time 0.04 seconds  
cpu time 0.03 seconds  
  
209  
210 DATA repo.dsense_6;  
211 set dsense_6;  
212 RUN;  
  
NOTE: There were 9 observations read from the data set WORK.DSENSE_6.  
NOTE: The data set REPO.DSENSE_6 has 9 observations and 18 variables.  
NOTE: DATA statement used (Total process time):  
real time 1.37 seconds  
cpu time 0.07 seconds  
  
213  
214  
215  
216 DATA dsense_9 (drop=Patient pat_id Time_to_treatment_month visit);  
217 set dsense;  
218 if visit="revisit" then delete;  
219 RUN;
```

Analysis and visualization

```
In [18]: #baseline  
dsense<-import("dsense_0.sas7bdat")  
dsense<-subset(dsense, select=c(Paper_code))  
names(dsense)[names(dsense) == "timepoint_HbA1c"] <- "HbA1c"  
  
# see ?cor for cor options. using pairwise.complete.obs because there are cases where all data for a v  
ariable is missing for a patient, but we dont want that entire patient point to be deleted  
#also decided to use kendalls given the small sample size, rank nature of method.  
m<-cor(dsense, use="pairwise.complete.obs", method="kendall")  
  
#see cor.test function for added parameters that can be used for cor.mtest  
#had to use exact=FALSE since there were ties and exact computations could not be performed  
#also as a result had to use continuity=TRUE since ties were present  
m.sig<-cor.mtest(dsense, method="kendall", conf.level=0.95, , continuity=TRUE, exact=FALSE)  
  
#for sig.level can introduce many different numbers. used 0 to hide the near perfect correlations (i.  
e. they wont receive a star)  
#order="AOE" (finding angular order of the eigenvectors) seems to be a good way to group vars related t  
o one other more closely together, but does not facilitate comparison between plots  
corrplot(m, method="circle",  
order="original",  
type="upper",  
tl.col="black",  
na.label = " ", # had to put this is because of the ? that results  
p.mat=m.sig$y, insig = "label_sig",sig.level = c(0, .05), pch.cex = 1.5, pch.col = "white")  
  
mtext("Correlation", side = 1, at=18.5, line=3.3, cex=.85)  
mtext("Coefficient", side = 1, at=18.5, line=4.0, cex=.85)
```



```
In [19]: #3 months  
dsense<-import("dsense_3.sas7bdat")  
dsense<-subset(dsense, select=c(Paper_code))  
names(dsense)[names(dsense) == "timepoint_HbA1c"] <- "HbA1c"  
  
# see ?cor for cor options. using pairwise.complete.obs because there are cases where all data for a v  
ariable is missing for a patient, but we dont want that entire patient point to be deleted  
#also decided to use kendalls given the small sample size, rank nature of method.  
m<-cor(dsense, use="pairwise.complete.obs", method="kendall")  
  
#see cor.test function for added parameters that can be used for cor.mtest  
#had to use exact=FALSE since there were ties and exact computations could not be performed  
#also as a result had to use continuity=TRUE since ties were present  
m.sig<-cor.mtest(dsense, method="kendall", conf.level=0.95, , continuity=TRUE, exact=FALSE)  
  
#for sig.level can introduce many different numbers. used 0 to hide the near perfect correlations (i.  
e. they wont receive a star)  
#order="AOE" (finding angular order of the eigenvectors) seems to be a good way to group vars related t  
o one other more closely together, but does not facilitate comparison between plots  
corrplot(m, method="circle",  
order="original",  
type="upper",  
tl.col="black",  
na.label = " ", # had to put this is because of the ? that results  
p.mat=m.sig$y, insig = "label_sig",sig.level = c(0, .05), pch.cex = 1.5, pch.col = "white")  
  
mtext("Correlation", side = 1, at=18.5, line=3.3, cex=.85)  
mtext("Coefficient", side = 1, at=18.5, line=4.0, cex=.85)
```



```
In [20]: #6 months  
dsense<-import("dsense_6.sas7bdat")  
dsense<-subset(dsense, select=c(Paper_code))  
names(dsense)[names(dsense) == "timepoint_HbA1c"] <- "HbA1c"  
  
# see ?cor for cor options. using pairwise.complete.obs because there are cases where all data for a v  
ariable is missing for a patient, but we dont want that entire patient point to be deleted  
#also decided to use kendalls given the small sample size, rank nature of method.  
m<-cor(dsense, use="pairwise.complete.obs", method="kendall")  
  
#see cor.test function for added parameters that can be used for cor.mtest  
#had to use exact=FALSE since there were ties and exact computations could not be performed  
#also as a result had to use continuity=TRUE since ties were present  
m.sig<-cor.mtest(dsense, method="kendall", conf.level=0.95, , continuity=TRUE, exact=FALSE)  
  
#for sig.level can introduce many different numbers. used 0 to hide the near perfect correlations (i.  
e. they wont receive a star)  
#order="AOE" (finding angular order of the eigenvectors) seems to be a good way to group vars related t  
o one other more closely together, but does not facilitate comparison between plots  
corrplot(m, method="circle",  
order="original",  
type="upper",  
tl.col="black",  
na.label = " ", # had to put this is because of the ? that results  
p.mat=m.sig$y, insig = "label_sig",sig.level = c(0, .05), pch.cex = 1.5, pch.col = "white")  
  
mtext("Correlation", side = 1, at=18.5, line=3.3, cex=.85)  
mtext("Coefficient", side = 1, at=18.5, line=4.0, cex=.85)
```



Heatmap showing the correlation of various clinical and immunological parameters with HbA1c. The color scale ranges from -1 (dark red) to 0.8 (dark blue).

Parameter	Correlation Coefficient
IFNg_C19A3	0.8
IL-10_C19A3	0.6
LST_CPM_PPI	0.4
IFNg_PPI	0.2
IL-10_PPI	0.0
Qdot_all PPI peptides	-0.2
Qdot_non-PPI	-0.4
LST_CPM_GAD65	-0.6
LST_CPM_IA-2	-0.8
LST_CPM_TT	-1.0
CD4 cl-4 Naïve	-0.8
CD4 cl-1 Trm	-0.6
CD4 cl-7 CM/EM	-0.4
CD4 subcl-Treg 6.4	-0.2
treatment_group	0.0
HbA1c	-1.0