Exploratory data analysis of genomic data from breast cancer studies

Urminder Singh

December 8, 2018

Introduction

Our data consists of gene expression data from various samples collected to study breast cancer (Breast Invasive Carcinoma). These samples were collected from a number of different individuals. Then, the samples' RNA were extracted sequenced to get RNA-seq reads from the sample. Using the RNA-seq, the gene expression was estimated. However, we used the data from (Wang et. al. 2018) where the expression data was mapped and normalized again to remove batch effects, which made the dataset comparable across the samples. Apart from the gene expression data for each sample, we also have associated metadata for each sample. This metadata provides information about the individual, tissue, disease etc.associated with each sample.

Our dataset has expression estimates of around 20,000 genes across 1,092 samples. Out of these 1,092 samples, 982 had tumor and 110 were normal tissues without tumor. We also added the gene metadata to our dataset which describes the gene features and functions. This also includes mutation information of the genes. This information can provide crucial information about how different mutations in the genome can lead to cancer.

Overall, our dataset is GxS matrix where each row corresponds to a gene and each column corresponds to a RNA-seq sample. Additionally, we have metadata for all the rows and all the columns.

Data

We used data from TCGA. TCGA is a comprehensive repository of human cancer molecular and clinical data, TCGA database has collected clinical and molecular phenotypes of thousands of tumor patients across different tumor types. The TCGA dataset, contains:

- Clinical information about participants
- Metadata about the samples (e.g. the weight of a sample portion, etc.)
- Histopathology slide images from sample portions
- Molecular information derived from the samples (e.g. mRNA/miRNA expression, protein expression, copy number, etc.)

First, we collected data clinical data from TCGA for breast cancer (BRCA) studies. This data contains Then we downloaded the mutation information for the BRCA studies. This data Finally, I collected the gene expression data from (Wang et. al. 2018). This data contains the estimated expression values of genes in the BRCA study. This data was normalized and batch corrected so it is comparable across different TCGA samples.

Data downloading and processing

This section describes how the data was downloaded and pre-processed for further analysis.

Downloading clinical data from TCGA

We used the TCGABiolinks package to download TCGA data. The function *GDCquery_clinic* is used to download the data. We wrote additional functions to clean the data and arrange it into a dataframe.

```
#required packages
library(TCGAbiolinks)
library(dplyr)
library(DT)
library(data.table)
library(plyr)
library(maftools)
library("readr")
library(ggplot2)

colsToKeep<-c("clinical.submitter_id","clinical.classification_of_tumor","clinical.primary_diagnosis","
#Function takes a df and expands it by unlisting elements at a column
expand<-function(df,colName){
  res<-data.frame()
  #for each row
  for(i in 1: dim(df)[1]){
    thisRow<-df[i, ! (colnames(df) %in% c(colName))]
    tempdf<-as.data.frame(df[i, c(colName)])
    #if list is empty skip that row
    if(dim(tempdf)[1]<1){
      next
    }
    #change colnames so they are unique
    colnames(tempdf)<-paste(paste(colName,".",sep = ""),colnames(tempdf),sep = "")
    #print(paste(i,colnames(tempdf)))
    newRow<-cbind(thisRow,tempdf,row.names = NULL)
    res<-bind_rows(res,newRow)

  }
  #print(res)
  return(res)
}

getjoinedBiospcClinc<-function(projName){
  print(paste("Downloading",projName))
  clinicalBRCA <- GDCquery_clinic(project = projName, type = "clinical")
  biospecimenBRCA <- GDCquery_clinic(project = projName, type = "Biospecimen")

  #rename all cols from clinical table with suffix clinical
  colnames(clinicalBRCA)<- paste0("clinical.",colnames(clinicalBRCA))

  #expand biospecimen data in the order portions, portions.analytes, portions.analytes.aliquots
  toUnpack<-c("portions", "portions.analytes", "portions.analytes.aliquots")
  for(s in toUnpack){
    biospecimenBRCA<-expand(biospecimenBRCA,s)
  }
  #add patient barcode to biospecimen data
  biospecimenBRCA<- biospecimenBRCA %>% mutate(clinical.bcr_patient_barcode=substr(submitter_id,1,nchar
  #join clinical and biospecimen
```

```
  finalJoined<-join(clinicalBRCA,biospecimenBRCA,by="clinical.bcr_patient_barcode")
  return(finalJoined)
}
########################################
##Download only BRCA metadata
brcaDF<-getjoinedBiospcClinc("TCGA-BRCA")
```

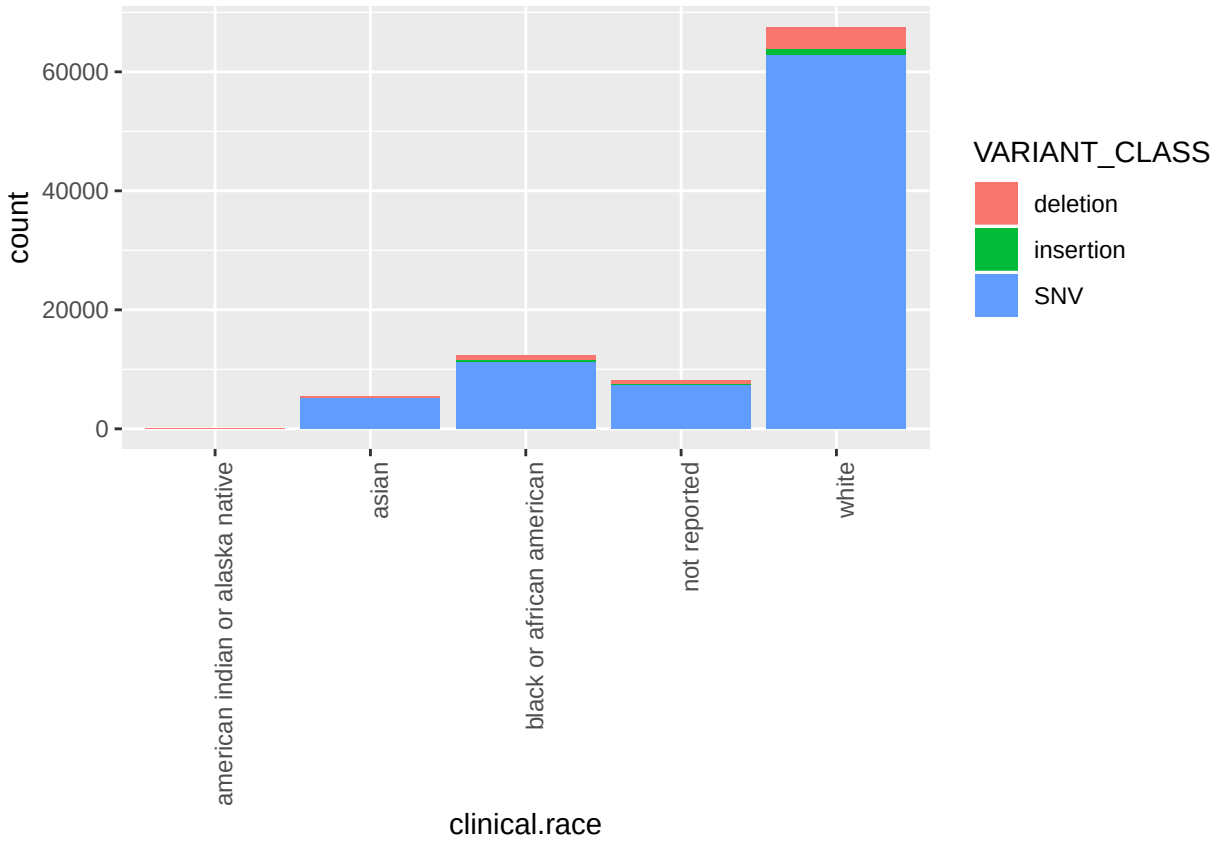
```
## [1] "Downloading TCGA-BRCA"
```

```
brcaDF<-brcaDF[,colsToKeep]
#remove cols with all NA values
naCols<-colnames(brcaDF)[sapply(brcaDF, function(x)all(is.na(x)))]
brcaDF<-brcaDF[,!(colnames(brcaDF) %in% naCols)]

head(brcaDF[,1:5],3)
```

```
##   clinical.submitter_id clinical.classification_of_tumor
## 1          TCGA-3C-AAAU                     not reported
## 2          TCGA-3C-AAAU                     not reported
## 3          TCGA-3C-AAAU                     not reported
##   clinical.primary_diagnosis clinical.tumor_stage
## 1     Lobular carcinoma, NOS              stage x
## 2     Lobular carcinoma, NOS              stage x
## 3     Lobular carcinoma, NOS              stage x
##   clinical.age_at_diagnosis
## 1                     20211
## 2                     20211
## 3                     20211
```

The clinical data for BRCA studies is stored in a dataframe *brcaDF*. Next, we downloaded mutation data using TCGABiolinks. After getting the mutation data, we joined the clinical table with the mutation table to have clinical information for each mutation in the mutation table.

```
brcaMAF <- GDCquery_Maf("BRCA", pipelines = "varscan2")
#for now read a copy which was already downloaded using above command

#join mutation information with the clinical data
colnames(brcaDF)[which(colnames(brcaDF)=="portions.analytes.aliquots.submitter_id")]="Tumor_Sample_Barc
brcaMAF_MD<-join(brcaMAF,brcaDF)
ggplot(data=brcaMAF_MD,aes(x=clinical.race,fill=VARIANT_CLASS))+geom_bar()+
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
```




```
head(brcaMAF,3)
head(brcaMAF_MD,3)
```

```
head(brcaMAF,3)
```

```
## # A tibble: 3 x 120
##   Hugo_Symbol Entrez_Gene_Id Center NCBI_Build Chromosome Start_Position
##   <chr>                <int> <chr>  <chr>      <chr>               <int>
## 1 CALML6              163688 WUGSC  GRCh38     chr1              1916819
## 2 PRKCZ                 5590 WUGSC  GRCh38     chr1              2172304
## 3 CCDC27              148870 WUGSC  GRCh38     chr1              3766586
## # ... with 114 more variables: End_Position <int>, Strand <chr>,
## #   Variant_Classification <chr>, Variant_Type <chr>,
## #   Reference_Allele <chr>, Tumor_Seq_Allele1 <chr>,
## #   Tumor_Seq_Allele2 <chr>, dbSNP_RS <chr>, dbSNP_Val_Status <chr>,
## #   Tumor_Sample_Barcode <chr>, Matched_Norm_Sample_Barcode <chr>,
## #   Match_Norm_Seq_Allele1 <lgl>, Match_Norm_Seq_Allele2 <lgl>,
## #   Tumor_Validation_Allele1 <lgl>, Tumor_Validation_Allele2 <lgl>,
## #   Match_Norm_Validation_Allele1 <lgl>,
## #   Match_Norm_Validation_Allele2 <lgl>, Verification_Status <lgl>,
## #   Validation_Status <lgl>, Mutation_Status <chr>,
## #   Sequencing_Phase <lgl>, Sequence_Source <lgl>,
## #   Validation_Method <lgl>, Score <lgl>, BAM_File <lgl>, Sequencer <chr>,
## #   Tumor_Sample_UUID <chr>, Matched_Norm_Sample_UUID <chr>, HGVSc <chr>,
## #   HGVSp <chr>, HGVSp_Short <chr>, Transcript_ID <chr>,
## #   Exon_Number <chr>, t_depth <int>, t_ref_count <int>,
## #   t_alt_count <int>, n_depth <int>, n_ref_count <lgl>,
```

```
## #   n_alt_count <lgl>, all_effects <chr>, Allele <chr>, Gene <chr>,
## #   Feature <chr>, Feature_type <chr>, One_Consequence <chr>,
## #   Consequence <chr>, cDNA_position <chr>, CDS_position <chr>,
## #   Protein_position <chr>, Amino_acids <chr>, Codons <chr>,
## #   Existing_variation <chr>, ALLELE_NUM <int>, DISTANCE <dbl>,
## #   TRANSCRIPT_STRAND <int>, SYMBOL <chr>, SYMBOL_SOURCE <chr>,
## #   HGNC_ID <chr>, BIOTYPE <chr>, CANONICAL <chr>, CCDS <chr>, ENSP <chr>,
## #   SWISSPROT <chr>, TREMBL <chr>, UNIPARC <chr>, RefSeq <chr>,
## #   SIFT <chr>, PolyPhen <chr>, EXON <chr>, INTRON <chr>, DOMAINS <chr>,
## #   GMAF <dbl>, AFR_MAF <dbl>, AMR_MAF <dbl>, ASN_MAF <lgl>,
## #   EAS_MAF <dbl>, EUR_MAF <dbl>, SAS_MAF <dbl>, AA_MAF <dbl>,
## #   EA_MAF <dbl>, CLIN_SIG <chr>, SOMATIC <lgl>, PUBMED <dbl>,
## #   MOTIF_NAME <lgl>, MOTIF_POS <lgl>, HIGH_INF_POS <lgl>,
## #   MOTIF_SCORE_CHANGE <lgl>, IMPACT <chr>, PICK <int>,
## #   VARIANT_CLASS <chr>, TSL <int>, HGVS_OFFSET <int>, PHENO <chr>,
## #   MINIMISED <int>, ExAC_AF <dbl>, ExAC_AF_Adj <dbl>, ExAC_AF_AFR <dbl>,
## #   ExAC_AF_AMR <dbl>, ExAC_AF_EAS <dbl>, ExAC_AF_FIN <dbl>, ...
```

```
head(brcaMAF_MD,3)
```

```
##   Hugo_Symbol Entrez_Gene_Id Center NCBI_Build Chromosome Start_Position
## 1      CALML6         163688  WUGSC     GRCh38       chr1        1916819
## 2       PRKCZ           5590  WUGSC     GRCh38       chr1        2172304
## 3      CCDC27         148870  WUGSC     GRCh38       chr1        3766586
##   End_Position Strand Variant_Classification Variant_Type Reference_Allele
## 1      1916819      +      Missense_Mutation          SNP                C
## 2      2172304      +      Missense_Mutation          SNP                G
## 3      3766586      +      Missense_Mutation          SNP                G
##   Tumor_Seq_Allele1 Tumor_Seq_Allele2 dbSNP_RS dbSNP_Val_Status
## 1                 C                 G     <NA>             <NA>
## 2                 G                 C     <NA>             <NA>
## 3                 G                 A     <NA>             <NA>
##           Tumor_Sample_Barcode  Matched_Norm_Sample_Barcode
## 1 TCGA-A2-A3Y0-01A-11D-A23C-09 TCGA-A2-A3Y0-10A-01D-A23C-09
## 2 TCGA-A2-A3Y0-01A-11D-A23C-09 TCGA-A2-A3Y0-10A-01D-A23C-09
## 3 TCGA-A2-A3Y0-01A-11D-A23C-09 TCGA-A2-A3Y0-10A-01D-A23C-09
##   Match_Norm_Seq_Allele1 Match_Norm_Seq_Allele2 Tumor_Validation_Allele1
## 1                     NA                     NA                       NA
## 2                     NA                     NA                       NA
## 3                     NA                     NA                       NA
##   Tumor_Validation_Allele2 Match_Norm_Validation_Allele1
## 1                       NA                           NA
## 2                       NA                           NA
## 3                       NA                           NA
##   Match_Norm_Validation_Allele2 Verification_Status Validation_Status
## 1                            NA                 NA                NA
## 2                            NA                 NA                NA
## 3                            NA                 NA                NA
##   Mutation_Status Sequencing_Phase Sequence_Source Validation_Method Score
## 1         Somatic               NA              NA                NA    NA
## 2         Somatic               NA              NA                NA    NA
## 3         Somatic               NA              NA                NA    NA
##   BAM_File           Sequencer                     Tumor_Sample_UUID
## 1       NA Illumina HiSeq 2000 d8fbb398-d1da-4444-984a-22c8523625da
## 2       NA Illumina HiSeq 2000 d8fbb398-d1da-4444-984a-22c8523625da
```

```
## 3       NA Illumina HiSeq 2000 d8fbb398-d1da-4444-984a-22c8523625da
##               Matched_Norm_Sample_UUID     HGVSc       HGVSp HGVSp_Short
## 1 46e1fd58-c2ea-46ef-bfc9-26d61e7be608  c.321C>G p.Asn107Lys     p.N107K
## 2 46e1fd58-c2ea-46ef-bfc9-26d61e7be608 c.1201G>C p.Gly401Arg     p.G401R
## 3 46e1fd58-c2ea-46ef-bfc9-26d61e7be608 c.1504G>A p.Glu502Lys     p.E502K
##     Transcript_ID Exon_Number t_depth t_ref_count t_alt_count n_depth
## 1 ENST00000307786         4/6      59          52           7      36
## 2 ENST00000378567       13/18      47          20          27      42
## 3 ENST00000294600        9/12      47          19          28      39
##   n_ref_count n_alt_count
## 1          NA          NA
## 2          NA          NA
## 3          NA          NA
##
## 1                                  CALML6,missense_variant,p.N107K,ENST00000307786,NM_138705.2,c.32
## 2 PRKCZ,missense_variant,p.G401R,ENST00000378567,NM_002744.4,c.1201G>C,MODERATE,YES,deleterious(0),p
## 3
##   Allele            Gene         Feature Feature_type  One_Consequence
## 1      G ENSG00000169885 ENST00000307786   Transcript missense_variant
## 2      C ENSG00000067606 ENST00000378567   Transcript missense_variant
## 3      A ENSG00000162592 ENST00000294600   Transcript missense_variant
##        Consequence cDNA_position CDS_position Protein_position Amino_acids
## 1 missense_variant      775/1101      321/546          107/181         N/K
## 2 missense_variant     1362/2326    1201/1779          401/592         G/R
## 3 missense_variant     1588/2176    1504/1971          502/656         E/K
##    Codons Existing_variation ALLELE_NUM DISTANCE TRANSCRIPT_STRAND SYMBOL
## 1 aaC/aaG               <NA>          1       NA                 1 CALML6
## 2 Ggc/Cgc               <NA>          1       NA                 1  PRKCZ
## 3 Gaa/Aaa               <NA>          1       NA                 1 CCDC27
##   SYMBOL_SOURCE    HGNC_ID        BIOTYPE CANONICAL        CCDS
## 1          HGNC HGNC:24193 protein_coding       YES CCDS30566.1
## 2          HGNC  HGNC:9412 protein_coding       YES    CCDS37.1
## 3          HGNC HGNC:26546 protein_coding       YES    CCDS50.1
##              ENSP SWISSPROT TREMBL       UNIPARC      RefSeq
## 1 ENSP00000304643    Q8TD86   <NA> UPI000034EC9B NM_138705.2
## 2 ENSP00000367830    Q05513   <NA> UPI0000169EB7 NM_002744.4
## 3 ENSP00000294600    Q2M243   <NA> UPI000013E186 NM_152492.2
##                SIFT                 PolyPhen  EXON INTRON
## 1 deleterious(0.01) probably_damaging(0.957)   4/6   <NA>
## 2    deleterious(0) probably_damaging(0.999) 13/18   <NA>
## 3    deleterious(0) probably_damaging(0.957)  9/12   <NA>
##
## 1
## 2 Pfam_domain:PF00069;Pfam_domain:PF07714;PROSITE_profiles:PS50011;SMART_domains:SM00220;SMART_domai
## 3
##   GMAF AFR_MAF AMR_MAF ASN_MAF EAS_MAF EUR_MAF SAS_MAF AA_MAF EA_MAF
## 1   NA      NA      NA      NA      NA      NA      NA     NA     NA
## 2   NA      NA      NA      NA      NA      NA      NA     NA     NA
## 3   NA      NA      NA      NA      NA      NA      NA     NA     NA
##   CLIN_SIG SOMATIC PUBMED MOTIF_NAME MOTIF_POS HIGH_INF_POS
## 1     <NA>      NA     NA         NA        NA           NA
## 2     <NA>      NA     NA         NA        NA           NA
## 3     <NA>      NA     NA         NA        NA           NA
##   MOTIF_SCORE_CHANGE   IMPACT PICK VARIANT_CLASS TSL HGVS_OFFSET PHENO
```

```
## 1                 NA MODERATE    1           SNV   1          NA  <NA>
## 2                 NA MODERATE    1           SNV   1          NA  <NA>
## 3                 NA MODERATE    1           SNV   1          NA  <NA>
##   MINIMISED ExAC_AF ExAC_AF_Adj ExAC_AF_AFR ExAC_AF_AMR ExAC_AF_EAS
## 1         1      NA          NA          NA          NA          NA
## 2         1      NA          NA          NA          NA          NA
## 3         1      NA          NA          NA          NA          NA
##   ExAC_AF_FIN ExAC_AF_NFE ExAC_AF_OTH ExAC_AF_SAS GENE_PHENO FILTER
## 1          NA          NA          NA          NA         NA   PASS
## 2          NA          NA          NA          NA         NA   PASS
## 3          NA          NA          NA          NA         NA   PASS
##       CONTEXT                            src_vcf_id
## 1 CAGAACCAGGA 83feabf8-563f-477c-b55c-40ab44223d1d
## 2 AGGAAGGCCTG 83feabf8-563f-477c-b55c-40ab44223d1d
## 3 TCATTGAAAAG 83feabf8-563f-477c-b55c-40ab44223d1d
##                         tumor_bam_uuid
## 1 73fc90f7-cb81-49c5-8677-3ad5bda5f7d9
## 2 73fc90f7-cb81-49c5-8677-3ad5bda5f7d9
## 3 73fc90f7-cb81-49c5-8677-3ad5bda5f7d9
##                        normal_bam_uuid
## 1 573e263e-39ad-4779-a7fd-5830684e9cb1
## 2 573e263e-39ad-4779-a7fd-5830684e9cb1
## 3 573e263e-39ad-4779-a7fd-5830684e9cb1
##                                case_id GDC_FILTER                  COSMIC
## 1 174850b4-5ec2-462b-a890-89bd1716b3c2       <NA>              COSM3803163
## 2 174850b4-5ec2-462b-a890-89bd1716b3c2       <NA> COSM3803781;COSM3803782
## 3 174850b4-5ec2-462b-a890-89bd1716b3c2       <NA>              COSM3805079
##   MC3_Overlap GDC_Validation_Status clinical.submitter_id
## 1        TRUE               Unknown          TCGA-A2-A3Y0
## 2        TRUE               Unknown          TCGA-A2-A3Y0
## 3        TRUE               Unknown          TCGA-A2-A3Y0
##   clinical.classification_of_tumor       clinical.primary_diagnosis
## 1                      not reported Infiltrating duct carcinoma, NOS
## 2                      not reported Infiltrating duct carcinoma, NOS
## 3                      not reported Infiltrating duct carcinoma, NOS
##   clinical.tumor_stage clinical.age_at_diagnosis clinical.vital_status
## 1            stage iib                     21024                 alive
## 2            stage iib                     21024                 alive
## 3            stage iib                     21024                 alive
##   clinical.days_to_death clinical.tissue_or_organ_of_origin
## 1                     NA                       Breast, NOS
## 2                     NA                       Breast, NOS
## 3                     NA                       Breast, NOS
##   clinical.days_to_birth clinical.site_of_resection_or_biopsy
## 1                 -21024                          Breast, NOS
## 2                 -21024                          Breast, NOS
## 3                 -21024                          Breast, NOS
##   clinical.days_to_last_follow_up clinical.gender clinical.year_of_birth
## 1                            1546          female                  1953
## 2                            1546          female                  1953
## 3                            1546          female                  1953
##               clinical.race     clinical.ethnicity clinical.year_of_death
## 1 black or african american not hispanic or latino                     NA
## 2 black or african american not hispanic or latino                     NA
```

```
## 3 black or african american not hispanic or latino                    NA
##   clinical.bcr_patient_barcode clinical.disease     submitter_id
## 1                 TCGA-A2-A3Y0             BRCA TCGA-A2-A3Y0-01A
## 2                 TCGA-A2-A3Y0             BRCA TCGA-A2-A3Y0-01A
## 3                 TCGA-A2-A3Y0             BRCA TCGA-A2-A3Y0-01A
##     sample_type portions.submitter_id portions.analytes.analyte_type
## 1 Primary Tumor   TCGA-A2-A3Y0-01A-11                            DNA
## 2 Primary Tumor   TCGA-A2-A3Y0-01A-11                            DNA
## 3 Primary Tumor   TCGA-A2-A3Y0-01A-11                            DNA
##   portions.analytes.submitter_id portions.analytes.analyte_type_id
## 1           TCGA-A2-A3Y0-01A-11D                                 D
## 2           TCGA-A2-A3Y0-01A-11D                                 D
## 3           TCGA-A2-A3Y0-01A-11D                                 D
##   portions.analytes.aliquots.analyte_type
## 1                                    <NA>
## 2                                    <NA>
## 3                                    <NA>
```

```
dim(brcaMAF)
```

```
## [1] 93612   120
```

```
dim(brcaMAF_MD)
```

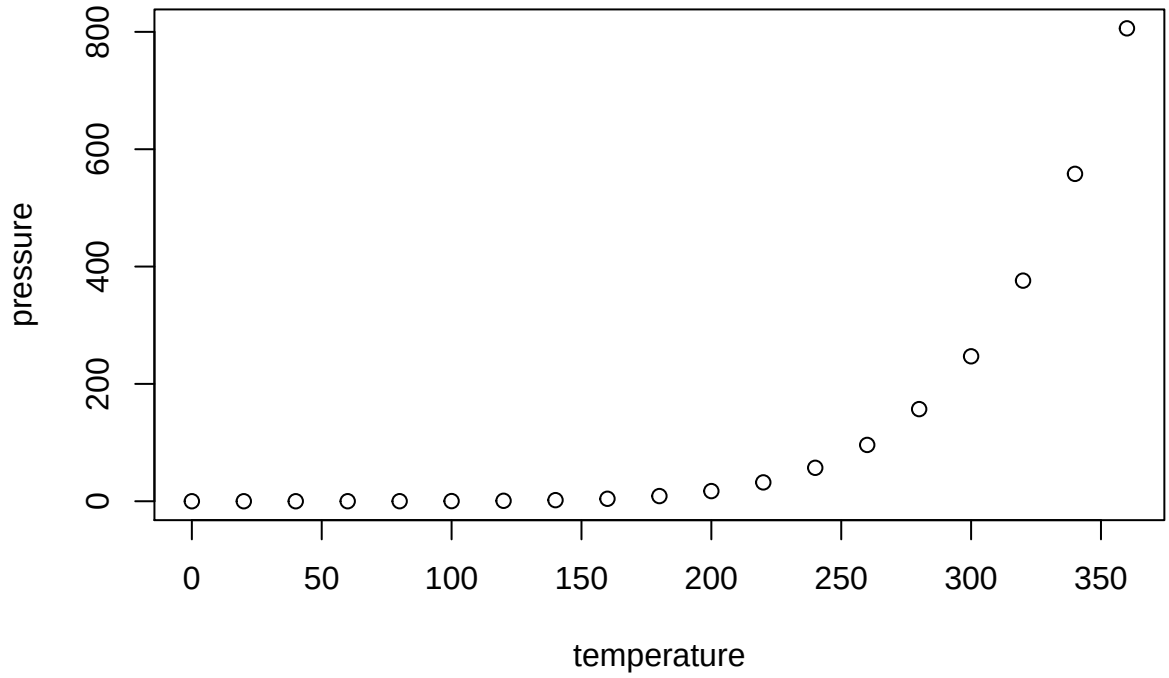
```
## [1] 93612   145
```

Processing

Analysis

You can also embed plots, for example:

```
plot(pressure)
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Conclusion

System information

```
sessionInfo()
```

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17134)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] parallel  stats     graphics  grDevices utils     datasets  methods
```

```
## [8] base
##
## other attached packages:
##  [1] ggplot2_3.1.0       readr_1.2.1         maftools_1.8.0
##  [4] Biobase_2.42.0      BiocGenerics_0.28.0 plyr_1.8.4
##  [7] data.table_1.11.8   DT_0.5              dplyr_0.7.8
## [10] TCGAbiolinks_2.10.0
##
## loaded via a namespace (and not attached):
##   [1] backports_1.1.2             circlize_0.4.5
##   [3] aroma.light_3.12.0          NMF_0.21.0
##   [5] selectr_0.4-1               ConsensusClusterPlus_1.46.0
##   [7] lazyeval_0.2.1              splines_3.5.1
##   [9] BiocParallel_1.16.2         GenomeInfoDb_1.18.1
##  [11] gridBase_0.4-7              sva_3.30.0
##  [13] digest_0.6.18               foreach_1.4.4
##  [15] htmltools_0.3.6             fansi_0.4.0
##  [17] magrittr_1.5                memoise_1.1.0
##  [19] BSgenome_1.50.0             cluster_2.0.7-1
##  [21] doParallel_1.0.14           limma_3.38.2
##  [23] ComplexHeatmap_1.20.0       Biostrings_2.50.1
##  [25] annotate_1.60.0             wordcloud_2.6
##  [27] matrixStats_0.54.0          R.utils_2.7.0
##  [29] prettyunits_1.0.2           colorspace_1.3-2
##  [31] blob_1.1.1                  rvest_0.3.2
##  [33] ggrepel_0.8.0               crayon_1.3.4
##  [35] RCurl_1.95-4.11             jsonlite_1.5
##  [37] genefilter_1.64.0           bindr_0.1.1
##  [39] survival_2.42-3             zoo_1.8-4
##  [41] iterators_1.0.10            glue_1.3.0
##  [43] survminer_0.4.3             registry_0.5
##  [45] gtable_0.2.0                zlibbioc_1.28.0
##  [47] XVector_0.22.0              GetoptLong_0.1.7
##  [49] DelayedArray_0.8.0          shape_1.4.4
##  [51] scales_1.0.0                DESeq_1.34.0
##  [53] rngtools_1.3.1              DBI_1.0.0
##  [55] edgeR_3.24.0                bibtex_0.4.2
##  [57] ggthemes_4.0.1              Rcpp_1.0.0
##  [59] xtable_1.8-3                progress_1.2.0
##  [61] cmprsk_2.2-7                mclust_5.4.2
##  [63] bit_1.1-14                  matlab_1.0.2
##  [65] km.ci_0.5-2                 stats4_3.5.1
##  [67] htmlwidgets_1.3             httr_1.3.1
##  [69] RColorBrewer_1.1-2          pkgconfig_2.0.2
##  [71] XML_3.98-1.16               R.methodsS3_1.7.1
##  [73] utf8_1.1.4                  locfit_1.5-9.1
##  [75] labeling_0.3                reshape2_1.4.3
##  [77] tidyselect_0.2.5            rlang_0.3.0.1
##  [79] AnnotationDbi_1.44.0        munsell_0.5.0
##  [81] tools_3.5.1                 cli_1.0.1
##  [83] downloader_0.4              RSQLite_2.1.1
##  [85] broom_0.5.0                 evaluate_0.12
##  [87] stringr_1.3.1               yaml_2.2.0
##  [89] knitr_1.20                  bit64_0.9-7
```

```
##  [91] survMisc_0.5.5             purrr_0.2.5
##  [93] bindrcpp_0.2.2             EDASeq_2.16.0
##  [95] nlme_3.1-137               formatR_1.5
##  [97] R.oo_1.22.0                xml2_1.2.0
##  [99] biomaRt_2.38.0             compiler_3.5.1
## [101] rstudioapi_0.8             curl_3.2
## [103] tibble_1.4.2               geneplotter_1.60.0
## [105] stringi_1.2.4              GenomicFeatures_1.34.1
## [107] lattice_0.20-35            Matrix_1.2-14
## [109] KMsurv_0.1-5               pillar_1.3.0
## [111] GlobalOptions_0.1.0        cowplot_0.9.3
## [113] bitops_1.0-6               rtracklayer_1.42.1
## [115] GenomicRanges_1.34.0       R6_2.3.0
## [117] latticeExtra_0.6-28        hwriter_1.3.2
## [119] ShortRead_1.40.0           gridExtra_2.3
## [121] IRanges_2.16.0             codetools_0.2-15
## [123] assertthat_0.2.0           SummarizedExperiment_1.12.0
## [125] pkgmaker_0.27              rprojroot_1.3-2
## [127] rjson_0.2.20               withr_2.1.2
## [129] GenomicAlignments_1.18.0   Rsamtools_1.34.0
## [131] S4Vectors_0.20.1           GenomeInfoDbData_1.2.0
## [133] mgcv_1.8-24                hms_0.4.2
## [135] grid_3.5.1                 tidyr_0.8.2
## [137] rmarkdown_1.10             ggpubr_0.2
```