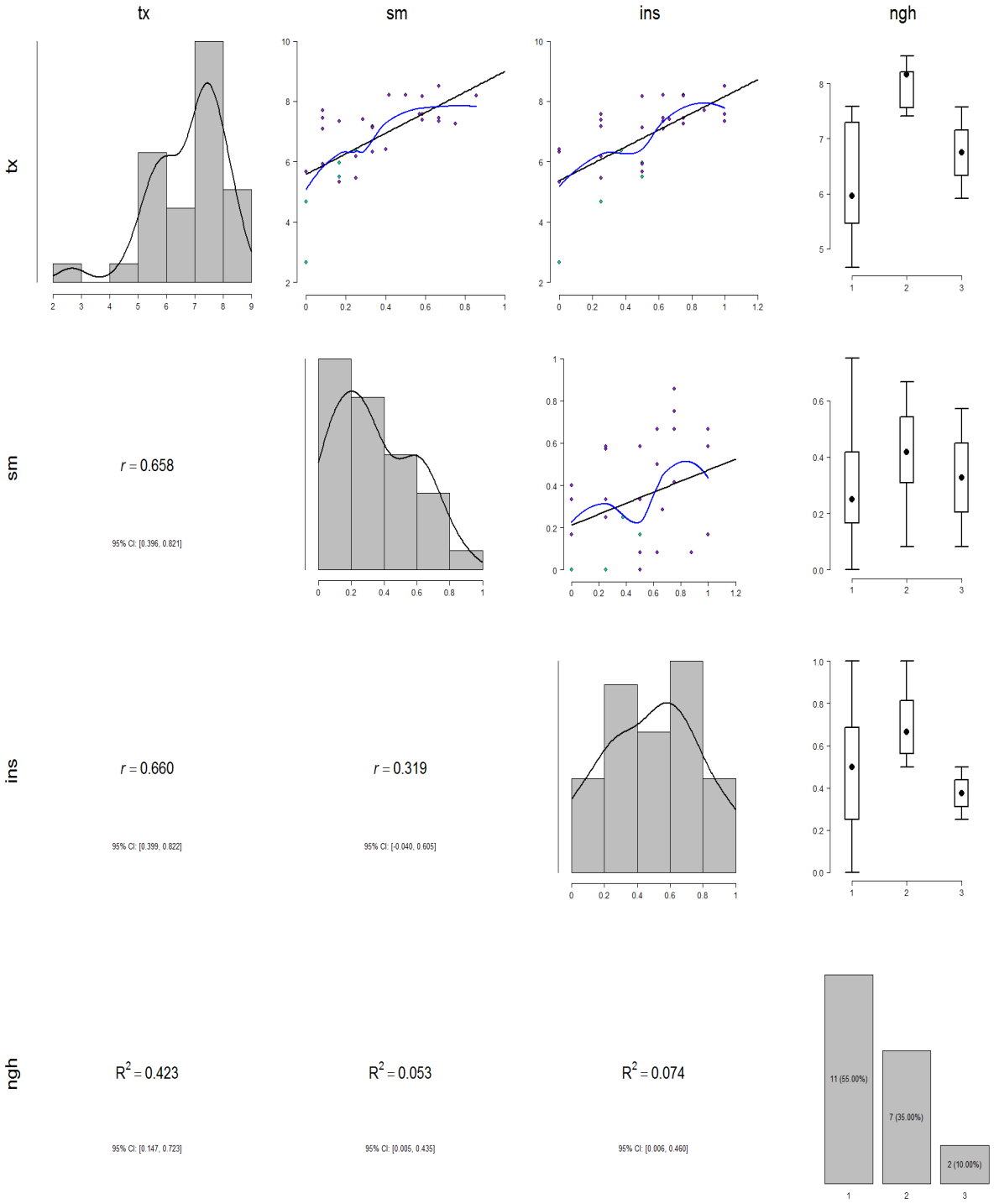
Appendix A.

Data Analysis from PhD Biostatistics Student



This figure shows marginal correlations between each variable (bottom left) along with scatter/box plots (top right) and univariate summaries (diagonal) for patients with SCD. Blue points are female, whereas purple points are male. Variables are:

1. tx = transition score
2. sm = self management score
3. ins = insurance score
4. ngh = number of guardians at home

First we use the data 'as-is', meaning we do not impute any missing values. We begin with a full model using all predictors

```
linmod = lm(tx~., data = mdat)

summary(linmod)

##
## Call:
## lm(formula = tx ~ ., data = mdat)
##
## Residuals:
##          1          2          4          5          6          7
## -2.599e-01 -2.111e-01 -1.569e-01 -2.597e-02 -2.493e-01  1.077e-01
##         10         11         12         13         14         15
##  1.643e-01 -2.426e-01  2.426e-01  2.479e-01  2.807e-01 -1.742e-01
##         19         20         21         24         26         27
##  3.469e-17  2.879e-01 -9.397e-02 -2.166e-01  3.311e-01  1.832e-01
##         28         31
##  3.301e-02 -2.479e-01
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.81966    1.96290   1.436   0.2104
## sm           1.76051    0.55706   3.160   0.0251 *
## tt           0.22606    1.80325   0.125   0.9051
## rx          -0.51352    1.27735  -0.402   0.7043
## adh          0.45311    0.49493   0.916   0.4019
## nt          -0.67037    0.94462  -0.710   0.5096
## ins          1.62792    0.44269   3.677   0.0143 *
## on1          0.45532    0.60713   0.750   0.4870
## new0.25      0.15182    0.54249   0.280   0.7908
## new0.5       0.34605    0.34612   1.000   0.3633
## new0.75      0.16911    0.50672   0.334   0.7521
## new1         0.68348    0.57469   1.189   0.2877
## ngh2         0.75304    0.49445   1.523   0.1883
## ngh3         0.49805    0.63021   0.790   0.4652
## age          0.13053    0.09332   1.399   0.2208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4173 on 5 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.966,  Adjusted R-squared:  0.8707
## F-statistic: 10.14 on 14 and 5 DF,  p-value: 0.009191
```

```
r2beta(linmod)
```

```
##     Effect   Rsq upper.CL lower.CL
## 1    Model 0.966    0.995    0.920
## 7      ins 0.730    0.952    0.328
## 2       sm 0.666    0.939    0.205
## 13    ngh2 0.317    0.850    0.002
## 15     age 0.281    0.838    0.002
## 12    new1 0.221    0.814    0.001
## 10  new0.5 0.167    0.789    0.001
## 5      adh 0.144    0.777    0.001
## 14    ngh3 0.111    0.758    0.000
## 8      on1 0.101    0.752    0.000
## 6       nt 0.092    0.746    0.000
## 4       rx 0.031    0.698    0.000
## 11 new0.75 0.022    0.689    0.000
## 9  new0.25 0.015    0.683    0.000
## 3       tt 0.003    0.670    0.000
```

Next, we select the most parsimonious model by taking out predictors (one at a time) that do not optimize Akaike's Information Criteria (AIC).

```
infmod = step(linmod, direction = 'backward', trace=0)
```

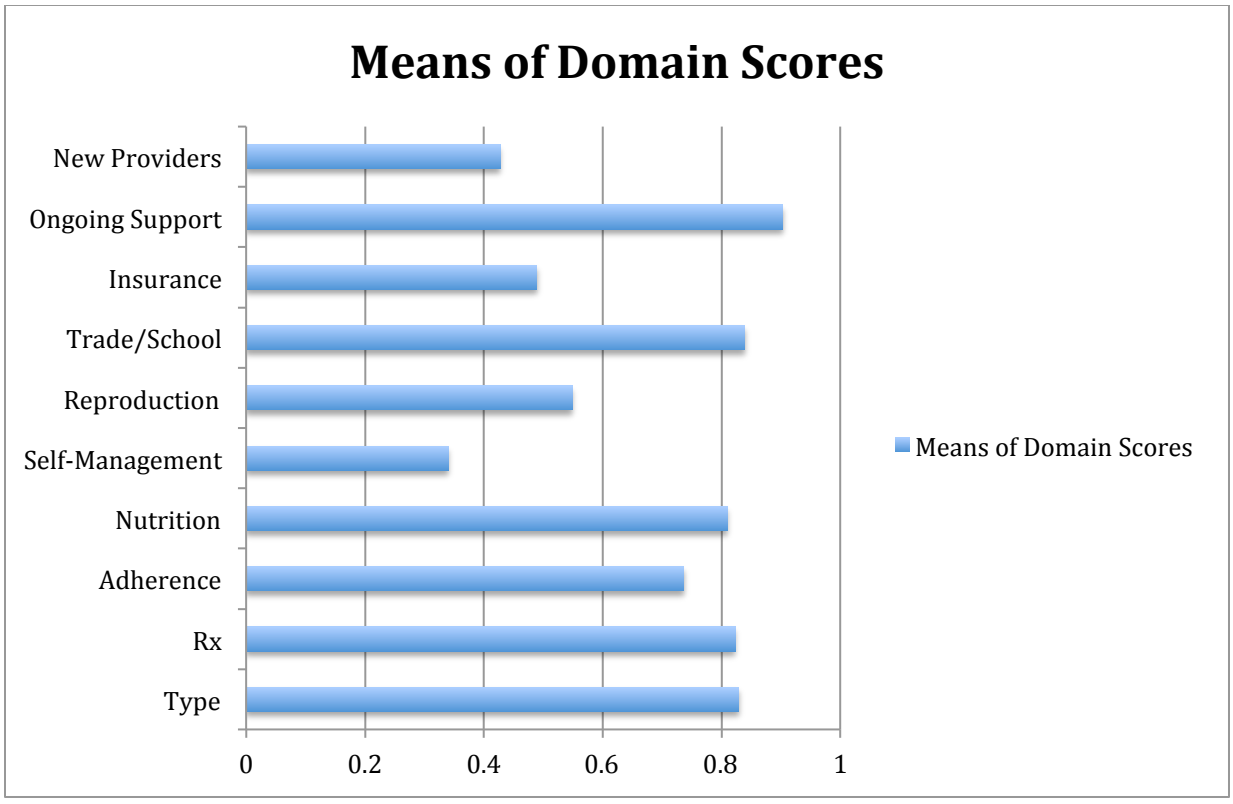
```
summary(infmod)
```

```
##
## Call:
## lm(formula = tx ~ sm + ins + on + ngh + age, data = mdat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44250 -0.20581 -0.05389  0.21961  0.52507
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.17835    0.77153   4.120 0.001208 **
## sm           1.84379    0.37651   4.897 0.000292 ***
## ins          1.65553    0.27102   6.108 3.73e-05 ***
## on1          0.44182    0.24665   1.791 0.096551 .
## ngh2         0.82674    0.20680   3.998 0.001519 **
## ngh3         0.51564    0.27121   1.901 0.079664 .
## age          0.08794    0.04806   1.830 0.090312 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.3366 on 13 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.9425, Adjusted R-squared:  0.9159
## F-statistic: 35.49 on 6 and 13 DF,  p-value: 2.505e-07

r2beta(infmod)

##   Effect   Rsq upper.CL lower.CL
## 1  Model 0.942    0.978    0.893
## 3    ins 0.742    0.897    0.518
## 2     sm 0.648    0.857    0.363
## 5   ngh2 0.551    0.812    0.224
## 6   ngh3 0.218    0.616    0.003
## 7    age 0.205    0.606    0.002
## 4    on1 0.198    0.601    0.002
```

11 of the participants had NA for number of guardians at home. We use multiple imputation with the mice package in R to imputing the 11 missing values for ngh. This changes the effect sizes slightly, but not enough to warrant any concern about the missing data.

```
## [1] "sm+tt+rx+adh+nt+ins+on+new+ngh+age"

##                est    se     t     df Pr(>|t|)  lo 95 hi 95 nmis   fmi
## (Intercept) 3.346 0.940 3.559 21.325    0.002  1.393 5.299   NA 0.118
## ins         1.805 0.379 4.767 20.593    0.000  1.017 2.594    0 0.148
## sm          1.878 0.495 3.798 21.419    0.001  0.851 2.905    0 0.114
## ngh2        0.709 0.285 2.490 20.471    0.021  0.116 1.303   NA 0.152
## ngh3        0.657 0.377 1.743 19.869    0.097 -0.130 1.443   NA 0.176
## age         0.063 0.063 1.005 21.475    0.326 -0.068 0.194    0 0.112
## on2         0.630 0.317 1.990 21.101    0.060 -0.028 1.288   NA 0.127
##             lambda
## (Intercept)  0.039
## ins          0.069
## sm           0.035
## ngh2         0.073
## ngh3         0.097
## age          0.032
## on2          0.048
```


Means of Domain Scores
New Providers
Ongoing Support
Insurance
Trade/School
Reproduction
Self-Management
Nutrition
Adherence
Rx
Type
0
0.2
0.4
0.6
0.8
1
Means of Domain Scores

Appendix B