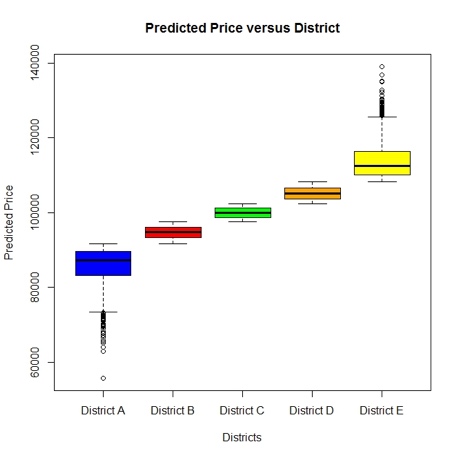
**Statistical Analysis Plan**

1. **Descriptive analysis**

You can then run t-tests and chi-square tests between this variable and other variables in the data set to see if the missingness on this variable is related to the values of other variables.

1. **Primary analysis**

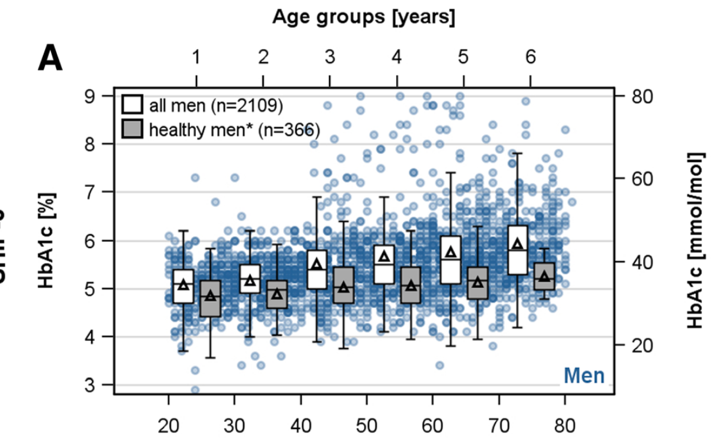
* Single linear regression (unadjusted)

1. hba1c = age[continuous]
2. hba1c = sex
3. 图表

   描述已自动生成hba1c = ethnicity (5)

* ANOVA test (multiple groups) & t-test (two groups) & multicollinearity test

\*Would also be helpful to check the interaction term.

1. hba1c with age[categorical] nested sex group
2. hba1c with age[categorical] nested ethnicity group
3. hba1c with sex nested ethnicity group
4. hba1c with pre- and post-menarche women nested ethnicity group 

* Multiple linear regression (adjusted)

1. hba1c = age[continuous] + sex + ethnicity (5), add interaction term based on the results from above.
2. 图表

   描述已自动生成linear regression assumption check
3. **Sensitivity analysis**

Include not known/not stated/missing ethnicity records. (Just in narrative maybe, any good idea?)

Remove records over 48 mmol/mol, check again. (because there might be patients with high hba1c but not recorded as diabetes petients)

1. **Issues may encounter**

What if not pass the multicollinearity test?

What if not pass the residual check?

What if sensitivity analysis shows totally different results (least likely happen)