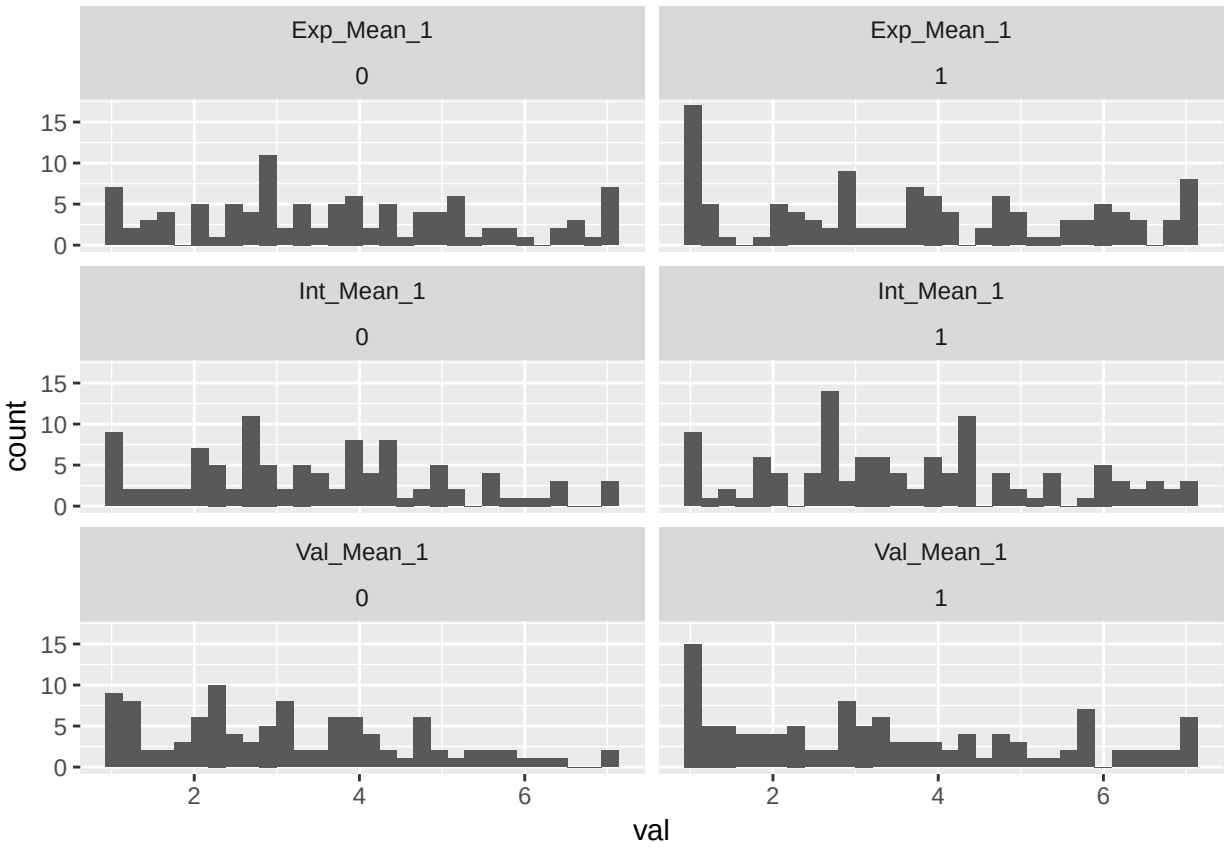
CT_Survey

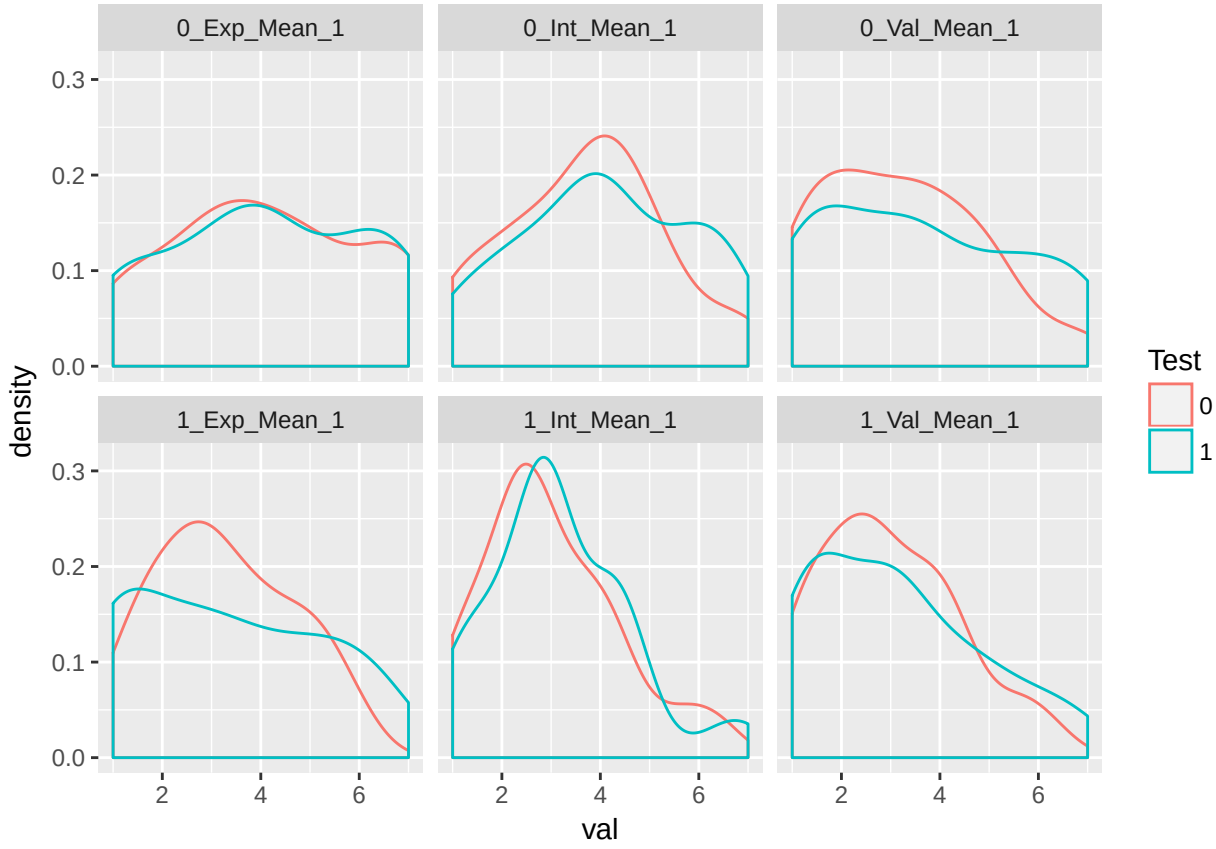
Mete and Josh

7/26/2017

```
library(tidyverse)
library(haven)

df <- read_sav("FullData.sav")

# df %>%
#   mutate(performance = ifelse(df$DBN_POMP == 0, 0, 1)) %>%
#   mutate(test_results = ifelse(DBN_POMP == 0, 0, 1)) %>%
#   group_by(ProgramingExtent, test_results) %>%
#   select(Exp_Mean_1, Int_Mean_1, Val_Mean_1) %>%
#   summarize_all(funs(mean)) %>%
#   mutate(test_results = as.factor(test_results))

df %>%
  select(contains("Mean_1"), Test) %>%
  gather(key, val, -Test) %>%
  mutate(Test = as.factor(Test)) %>%
  ggplot(aes(x = val)) +
  geom_histogram() +
  facet_wrap( ~ key + Test, ncol = 2)
```



```
df %>%
  select(contains("Mean_1"), Test, Gender) %>%
  gather(key, val, -Test, -Gender) %>%
  mutate(Test = as.factor(Test)) %>%
  filter(!is.nan(Gender)) %>%
  unite(facet_var, Gender, key) %>%
  ggplot(aes(x = val, colour = Test)) +
  geom_density() +
  facet_wrap( ~ facet_var)
```




```
# girls possibly calibrate their expectancies
# boys possibly get more interested
# girls and boys possibly both value it more

df$DBN_POMP
```

```
##   [1] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN
##  [18] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN
##  [35] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN
##  [52] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN
##  [69] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN
##  [86] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN
## [103] NaN   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## [120]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## [137]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## [154]   0   0   0  17  17  17  17  17  17  17  17  17  17  17  17  17  17
## [171]  17  17  17  17  17  17  17  17  17  17  17  17  17  33  33  33  33
```

```
## [188]  33  33  33  33  33  33  33  33  33  33  33  33  33  50  50  50  50
## [205]  50  50  50  50  50  50  67  67  67  67  67  83
## attr(,"format.spss")
## [1] "F2.1"
```

```
df %>%
  select(contains("Mean_1"), Test, Gender, DBN_POMP) %>%
  mutate(DBN_POMP_0 = ifelse(DBN_POMP == 0, 0, 1),
         DBN_POMP_17 = ifelse(DBN_POMP <= 17, 0, 1)) %>%
  filter(!is.nan(Gender)) %>%
  group_by(Test, Gender, DBN_POMP_0)  %>%
  summarize_if(is.numeric, funs(mean)) %>%
  select(Test, Gender, DBN_POMP_0, everything(), -DBN_POMP, -DBN_POMP_17)
```

```
## # A tibble: 6 x 6
## # Groups:   Test, Gender [4]
##        Test    Gender DBN_POMP_0 Int_Mean_1 Val_Mean_1 Exp_Mean_1
##   <dbl+lbl> <dbl+lbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1         0         0         NA   3.710345   3.234553   4.128211
## 2         0         1         NA   3.006977   2.983721   3.255814
## 3         1         0          0   3.537668   2.946802   3.748624
## 4         1         0          1   4.603030   4.333333   4.451098
## 5         1         1          0   3.156522   3.430435   3.608696
## 6         1         1          1   3.176923   2.857692   3.087932
```

```
df <- df %>%
  mutate(DBN_POMP_0 = ifelse(DBN_POMP == 0, 0, 1),
         DBN_POMP_17 = ifelse(DBN_POMP <= 17, 0, 1),
         Test = as.factor(Test),
         DBN_POMP_0 = as.factor(DBN_POMP_0))

df_ss <- filter(df, !is.na(DBN_POMP_0))
df_ss <- filter(df_ss, !is.nan(Gender))

df_ss %>% select(Int_Mean_1, Test, Gender, DBN_POMP_0) %>%
  count(Gender)
```

```
## # A tibble: 2 x 2
##      Gender     n
##   <dbl+lbl> <int>
## 1         0    63
## 2         1    49
```

```
m1 <- aov(Int_Mean_1 ~ as.factor(Gender) * DBN_POMP_0, data = df_ss)
summary(m1)
```

```
##                               Df Sum Sq Mean Sq F value  Pr(>F)
## as.factor(Gender)              1  23.76  23.755   9.561 0.00253 **
## DBN_POMP_0                     1  10.34  10.340   4.162 0.04379 *
## as.factor(Gender):DBN_POMP_0   1   7.50   7.501   3.019 0.08515 .
## Residuals                    108 268.34   2.485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(m1)
```

```
##   Tukey multiple comparisons of means
```

```
##     95% family-wise confidence level
##
## Fit: aov(formula = Int_Mean_1 ~ as.factor(Gender) * DBN_POMP_0, data = df_ss)
##
## $`as.factor(Gender)`
##           diff       lwr        upr     p adj
## 1-0 -0.9283681 -1.523501 -0.3332354 0.0025296
##
## $DBN_POMP_0
##          diff        lwr     upr     p adj
## 1-0 0.6085459 0.01723195 1.19986 0.0437968
##
## $`as.factor(Gender):DBN_POMP_0`
##                 diff         lwr        upr     p adj
## 1:0-0:0 -0.38114642 -1.52113248  0.7588396 0.8190691
## 0:1-0:0  1.06536214  0.02774096  2.1029833 0.0418240
## 1:1-0:0 -0.36074508 -1.46287519  0.7413850 0.8282951
## 0:1-1:0  1.44650856  0.32923563  2.5637815 0.0055244
## 1:1-1:0  0.02040134 -1.15702460  1.1978273 0.9999666
## 1:1-0:1 -1.42610723 -2.50472734 -0.3474871 0.0043888
```

```
m2 <- aov(Val_Mean_1 ~ as.factor(Gender) * DBN_POMP_0, data = df_ss)
summary(m2)
```

```
##                               Df Sum Sq Mean Sq F value  Pr(>F)
## as.factor(Gender)              1    8.2   8.233   2.517 0.11558
## DBN_POMP_0                     1    7.8   7.844   2.398 0.12445
## as.factor(Gender):DBN_POMP_0   1   26.4  26.369   8.060 0.00541 **
## Residuals                    108  353.3   3.272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(m2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Val_Mean_1 ~ as.factor(Gender) * DBN_POMP_0, data = df_ss)
##
## $`as.factor(Gender)`
##           diff       lwr       upr     p adj
## 1-0 -0.5465496 -1.229467 0.1363679 0.1155803
##
## $DBN_POMP_0
##         diff        lwr      upr     p adj
## 1-0 0.530041 -0.1484945 1.208577 0.1244554
##
## $`as.factor(Gender):DBN_POMP_0`
##                 diff        lwr        upr     p adj
## 1:0-0:0  0.48363298 -0.8245064  1.7917724 0.7697002
## 0:1-0:0  1.38653153  0.1958563  2.5772068 0.0155516
## 1:1-0:0 -0.08910949 -1.3538090  1.1755900 0.9977812
## 0:1-1:0  0.90289855 -0.3791774  2.1849745 0.2614879
## 1:1-1:0 -0.57274247 -1.9238443  0.7783594 0.6865207
## 1:1-0:1 -1.47564103 -2.7133627 -0.2379193 0.0125729
```

```r
m3 <- aov(Exp_Mean_1 ~ as.factor(Gender) * DBN_POMP_0, data = df_ss)
summary(m3)
```

```
##                               Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(Gender)              1   17.0  16.951   4.541 0.0354 *
## DBN_POMP_0                     1    0.8   0.786   0.210 0.6473
## as.factor(Gender):DBN_POMP_0   1   10.3  10.279   2.753 0.1000 .
## Residuals                    108  403.2   3.733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(m3)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Exp_Mean_1 ~ as.factor(Gender) * DBN_POMP_0, data = df_ss)
##
## $`as.factor(Gender)`
##           diff       lwr         upr     p adj
## 1-0 -0.7842144 -1.513697 -0.05473137 0.0353671
##
## $DBN_POMP_0
##          diff       lwr       upr     p adj
## 1-0 0.1677472 -0.557055 0.8925495 0.6473345
##
## $`as.factor(Gender):DBN_POMP_0`
##                diff        lwr         upr     p adj
## 1:0-0:0 -0.1399282 -1.5372645  1.25740812 0.9937138
## 0:1-0:0  0.7024743 -0.5693884  1.97433700 0.4766290
## 1:1-0:0 -0.6606916 -2.0116260  0.69024283 0.5800064
## 0:1-1:0  0.8424025 -0.5270932  2.21189822 0.3800740
## 1:1-1:0 -0.5207634 -1.9639916  0.92246480 0.7825221
## 1:1-0:1 -1.3631659 -2.6852830 -0.04104879 0.0406194
```