1) Multivariate analysis is the statistical study where there is more than one outcome variable.

One research paper with multivariate analysis as its focus is the paper on MANOVA(below is the reference).

French, A., Macedo, M., Poulsen, J., Waterson, T. and Yu, A.

Multivariate Analysis of Variance (MANOVA)

In-text: (French et al, n.d.)

Bibliography: French, A., Macedo, M., Poulsen, J., Waterson, T. and Yu, A., n.d. *Multivariate Analysis of Variance (MANOVA)*. 1st ed.

For this paper the independent variables are several different textbooks. While the dependent variables are improvement in math and physics skills.

2)

Within the Olympics dataset there are a few interesting tends. The most interesting trend is that the GDP of a country has a huge correcaltion with the number of metals it recevieces. However, lesser metal rankings show a lessor degree of statistical significance. That said GDP is statically significant for each metal.

```
Call:
lm(formula = olympics$Gold.medals ~ olympics$X2011.GDP)

Residuals:
    Min      1Q  Median      3Q     Max
-13.548  -2.384  -1.869   0.444  19.114

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.340e+00  2.110e+00   1.109    0.282
olympics$X2011.GDP 3.103e-12  4.975e-13   6.238 6.95e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.963 on 18 degrees of freedom
Multiple R-squared:  0.6837,   Adjusted R-squared:  0.6661
F-statistic: 38.91 on 1 and 18 DF,  p-value: 6.951e-06


Call:
lm(formula = olympics$Silver.medals ~ olympics$X2011.GDP)

Residuals:
    Min      1Q  Median      3Q     Max
-7.2284 -3.5947 -2.4296  0.8607 18.4024

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)        3.579e+00  1.665e+00   2.150   0.0454 *
olympics$X2011.GDP 2.163e-12  3.925e-13   5.511 3.12e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.282 on 18 degrees of freedom
Multiple R-squared:  0.6279,   Adjusted R-squared:  0.6072
F-statistic: 30.37 on 1 and 18 DF,  p-value: 3.115e-05


Call:
lm(formula = olympics$Bronze.medals ~ olympics$X2011.GDP)

Residuals:
   Min     1Q Median     3Q    Max
-6.513 -3.910 -1.730  1.679 23.501

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        4.707e+00  1.803e+00   2.611 0.017696 *
olympics$X2011.GDP 2.041e-12  4.251e-13   4.801 0.000143 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.805 on 18 degrees of freedom
Multiple R-squared:  0.5615,   Adjusted R-squared:  0.5371
F-statistic: 23.05 on 1 and 18 DF,  p-value: 0.0001432
```

If you look at figure #1 one you can see the relationship that GDP has to the number of metals won. Thus, how much overall a country can produce effects how well it does in the Olympics. The few exceptions seem to be RUS, GBR along with CHN to some extent. These countries have a large gap between the number of metals above the GDP compared to the peers. This surplus gap is pretty large with these three countries.

Looking at the other variable of the country itself is the population. If you look at figure #2 it shows the population compared to the number of medals. However, there seems to be little correlation first shown by the graph and also show if an regression model is made using population as a single variable. Even taking an interaction between the two and looking at the GDP per capita also reflects no real relationship. This can be shown in figure #3.

If you focus on just the number of contestants(both male and female) the countries bring you get these results from a regression model for each medal.

```
Call:
lm(formula = olympics$Gold.medals ~ (olympics$Male.count +
olympics$Female.count))

Residuals:
    Min      1Q  Median      3Q     Max
-14.322  -1.407   0.141   2.027  13.154

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          -0.77706    2.01000  -0.387 0.703850
olympics$Male.count  -0.13866    0.06111  -2.269 0.036559 *
olympics$Female.count  0.25711    0.06026   4.266 0.000521 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.094 on 17 degrees of freedom
Multiple R-squared:  0.8251,   Adjusted R-squared:  0.8045
F-statistic: 40.09 on 2 and 17 DF,  p-value: 3.668e-07


Call:
lm(formula = olympics$Silver.medals ~ (olympics$Male.count + 
    olympics$Female.count))

Residuals:
    Min      1Q  Median      3Q     Max
-7.2513 -0.9472 -0.1289  2.0810  4.8901

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.14070    1.19292   0.118 0.907496
olympics$Male.count   -0.05645    0.03627  -1.557 0.137975
olympics$Female.count  0.14986    0.03577   4.190 0.000615 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.617 on 17 degrees of freedom
Multiple R-squared:  0.8835,   Adjusted R-squared:  0.8698
F-statistic: 64.47 on 2 and 17 DF,  p-value: 1.157e-08


Call:
lm(formula = olympics$Bronze.medals ~ (olympics$Male.count + 
    olympics$Female.count))

Residuals:
    Min      1Q  Median      3Q     Max
-6.1086 -1.3691 -0.0841  1.1646  8.4932

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.98739    1.06869   0.924 0.368456
olympics$Male.count   -0.06330    0.03249  -1.948 0.068069 .
olympics$Female.count  0.15721    0.03204   4.907 0.000133 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.24 on 17 degrees of freedom
Multiple R-squared:  0.9061,   Adjusted R-squared:  0.8951
F-statistic: 82.02 on 2 and 17 DF,  p-value: 1.852e-09
```

As you can see overall the model is statistical significance for each medal. What is interesting in these results is that the number of female contestants shows to be more useful to the model than the number of male contestants. If you look at figure #4 you can see a graph showing the total number of contestants to each medal by country. The outliers here that seem to do well against their peers would be CHN and a little bit USA. Once again judging by the gap between the black line(total number of contestants) and the number of medals. The countries that seemed to do poor would be GBR,DEU,BRA,FRA,ITA.

3)

The canonical correlation analysis(CCA) involves several metric dependent variables and several metric independent variables. The dependent variables in this case are sales growth, sales profitability and new account sales. The independent variables in this case are attributes of the employees in the area of creativity, mechanical reasoning, abstract reasoning and mathematics. The CCA in this case will have three functions since that is the lowest of the number of variables from the dependent and independent variables. Now each of these functions will represent capturing some different underlying information with the result being to maximize the correlation between the two sets. So the loadings that the analysis uses can differ from each other and send a mix message but this is OK. It's OK since they are all orthogonal to each other thus each are just capturing different information.

The functions scores for the three functions are as followed: 0.99 , 0.88 , 0.38 . Now the highest possible score would be the number 1 while the lowest being 0. As you can see the first two functions are very high. Given the information inferred from the loadings function one will have a slight more importance than function two. However, if you look at the third function it is only .38 less than half of the second function. While it is low the information still can be use to make inferences but it should be noticed that the information here is much less important than the first two. On a final note the inferences do not mean a cause and effect it will just show that they have a relationship to each other.

First Function of CCA:

For the first function of CCA model all of the dependent variables: growth, profit and new accounts show a very high positive relationship with sales. What is interesting in this model is that they are all clustered very close together showing they have about the same level of importance. While on the independent variable side the mathematics ability seems a much greater importance followed by mechanical reasoning. While the creativity and abstract reasoning skills seem to be about tied but much lower than mathematics, the highest score . It should also be noted that the relationships are all positive. Overall for the first function it has a canonical correlation of .99 thus about 99% of the underlining variance of the data being used for THIS function is explained by the variance in the employees skills.

Second Function of CCA:

Now what is really interesting with the second model is the spread in the dependent loading variables scores. Model one had a spread of .03 while this one has a spread of .51 . In this function the highest influencing variable is profit with a positive correlation. The dependent growth seems to be neutral in this function. Now, new accounts in this model has a negative value compared to the positive in model one. Moving on to the independent loading variable scores. Abstract reasoning now is the biggest influence with a negative relationship. Creativity also now has a negative relationship. It is interesting to note that mechanical reasoning and Math are still positive. This is interesting because they were noticeably higher positive values on model one. Overall for the second function it has a canonical

correlation of .88 thus about 88% of the underlining variance of the data being used for THIS function is explained by the variance in the employees skills.

Third Function of CCA:

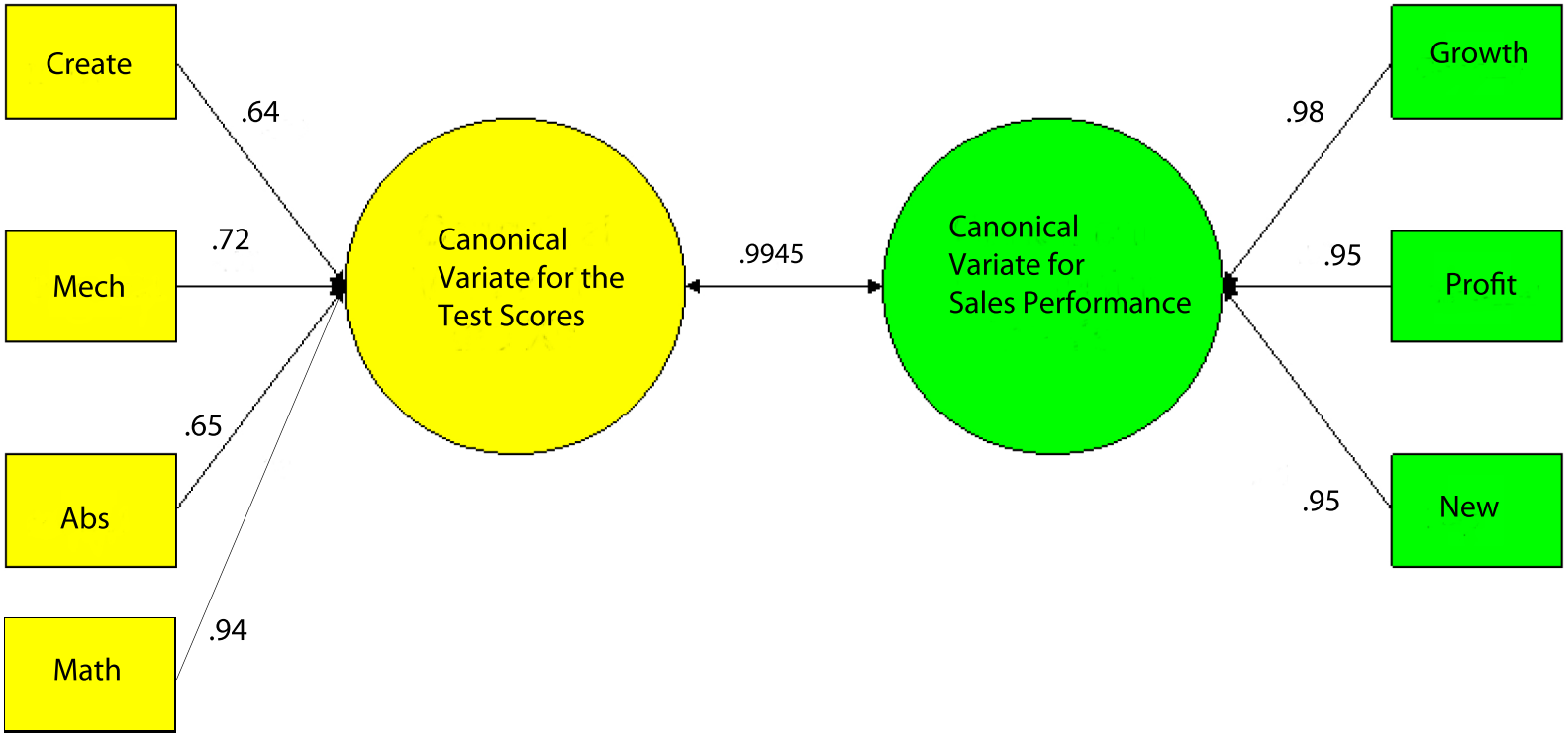
For the dependent loading variables scores on the third model the scores are quite low. The range is still about the same as function two at .44 . However, it is important to note that the profit variable unlike the last two models is not positive it is now neutral. While new accounts has changed. It about the same score however now its negative compared to the last two models of being positive. As for the independent loading variables the results are very different than the first two. While mechanical reasoning and mathematical skills have been cluster close together in the past two with having a large positive value. Now they are closer to neutral but being a little navigate. On the flip side creativity and abstract reasoning are no longer cluster together like they have been in the past. Creativity is highly positive and abstracting reasoning is highly negative.

Overall brake down from using information from the three functions .

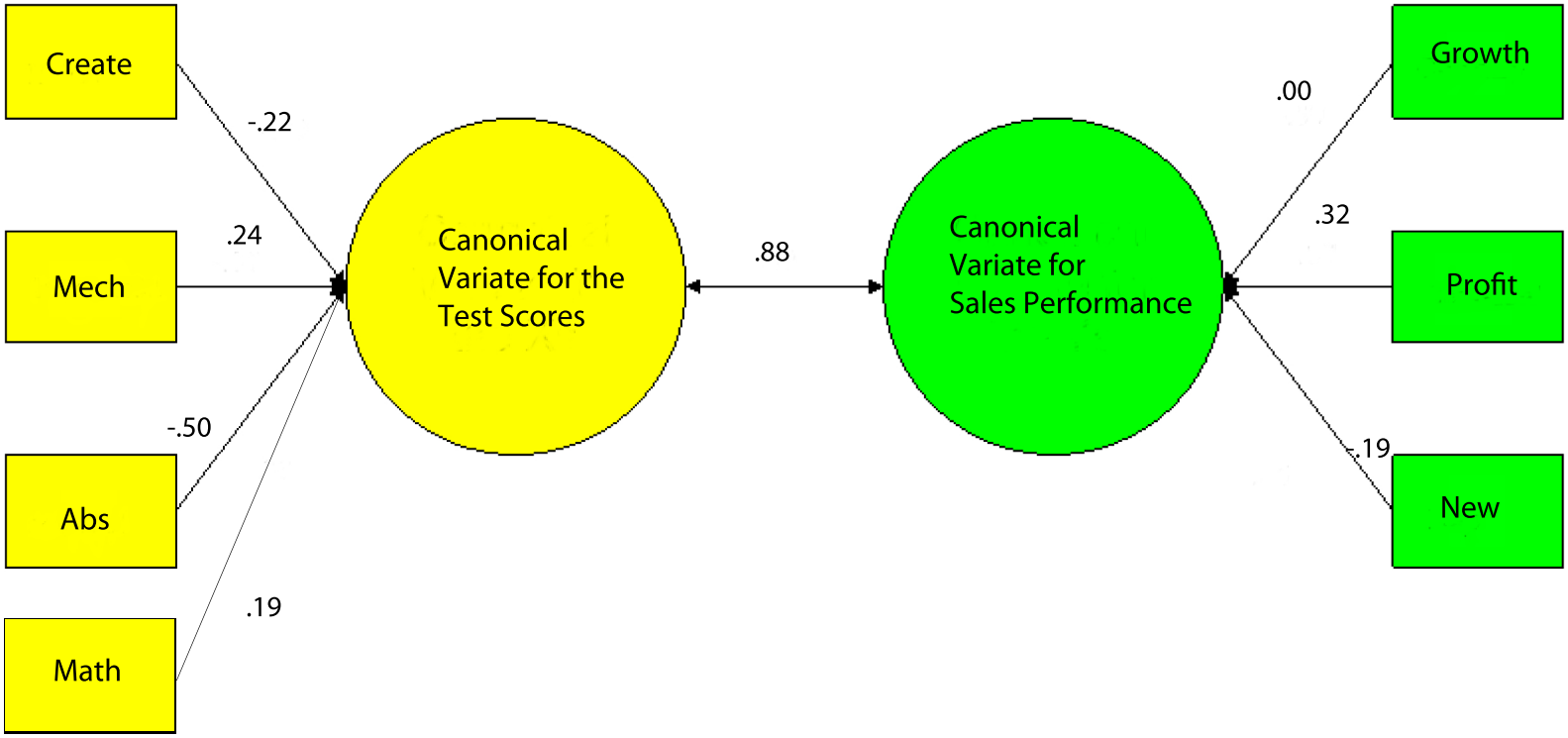
Dependence variables order of importance is profit, growth, new accounts.

For the independent variables order of importance is math, mechanical reasoning, creativity and abstract thinking. It should be noted that the first two are much more important than the last two.

Frist Function of CCA



Second Function of CCA



Third Function of CCA

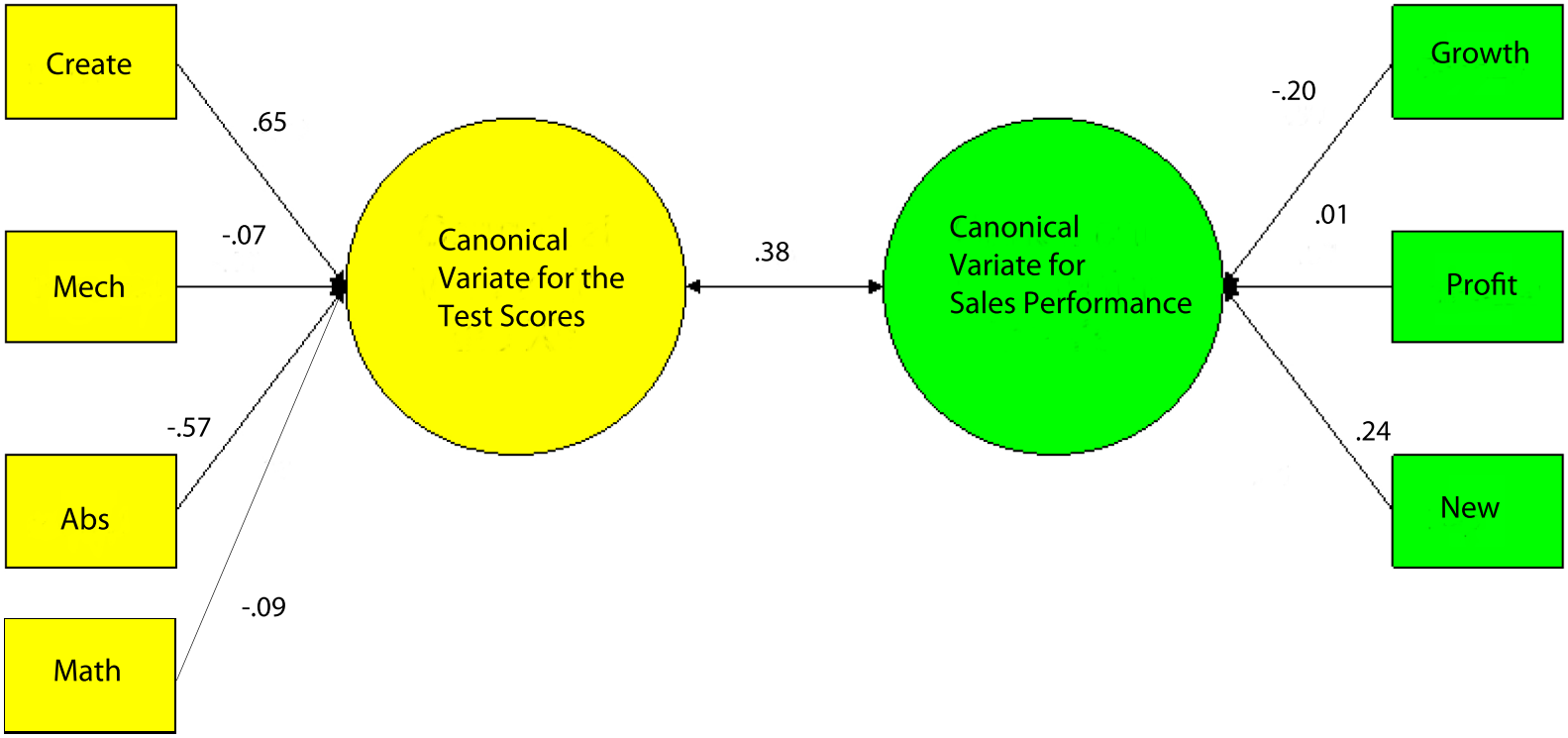


Figure 2

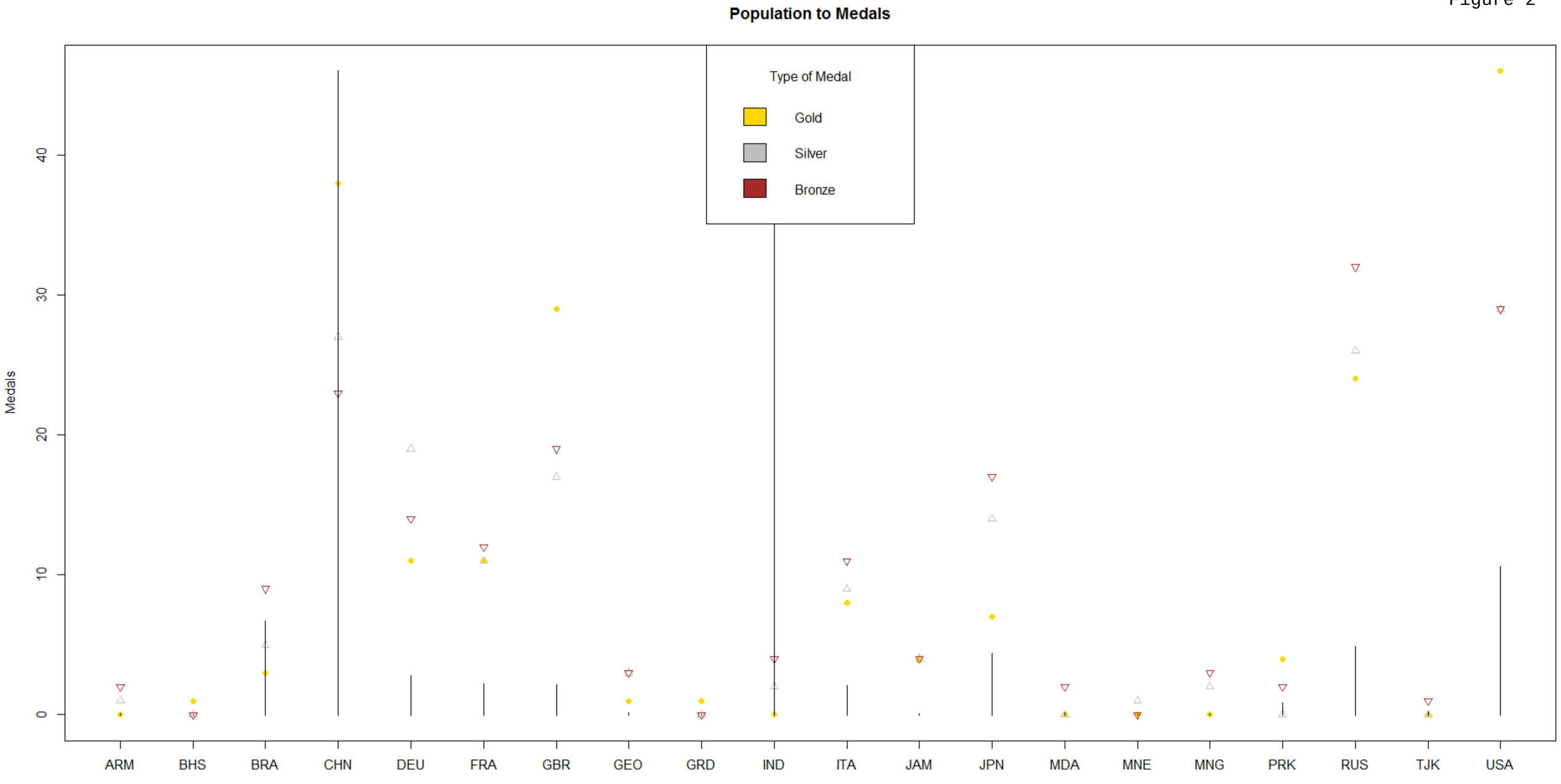

Population to Medals
Type of Medal
Gold
Silver
Bronze
Medals
0
10
20
30
40
ARM
BHS
BRA
CHN
DEU
FRA
GBR
GEO
GRD
IND
ITA
JAM
JPN
MDA
MNE
MNG
PRK
RUS
TJK
USA

Figure 3

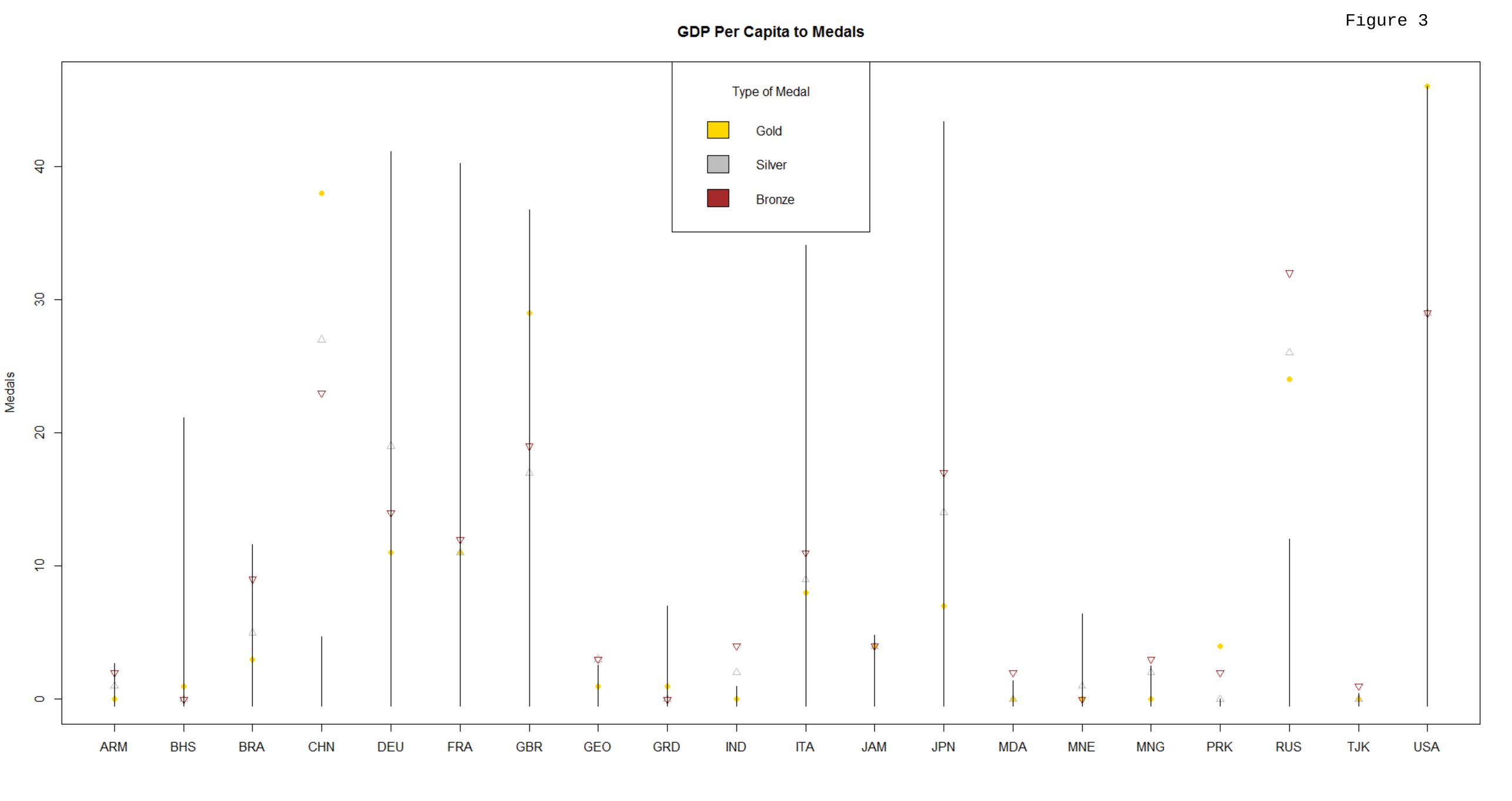

GDP Per Capita to Medals
Type of Medal
Gold
Silver
Bronze
Medals
0
10
20
30
40
ARM
BHS
BRA
CHN
DEU
FRA
GBR
GEO
GRD
IND
ITA
JAM
JPN
MDA
MNE
MNG
PRK
RUS
TJK
USA

Figure 1

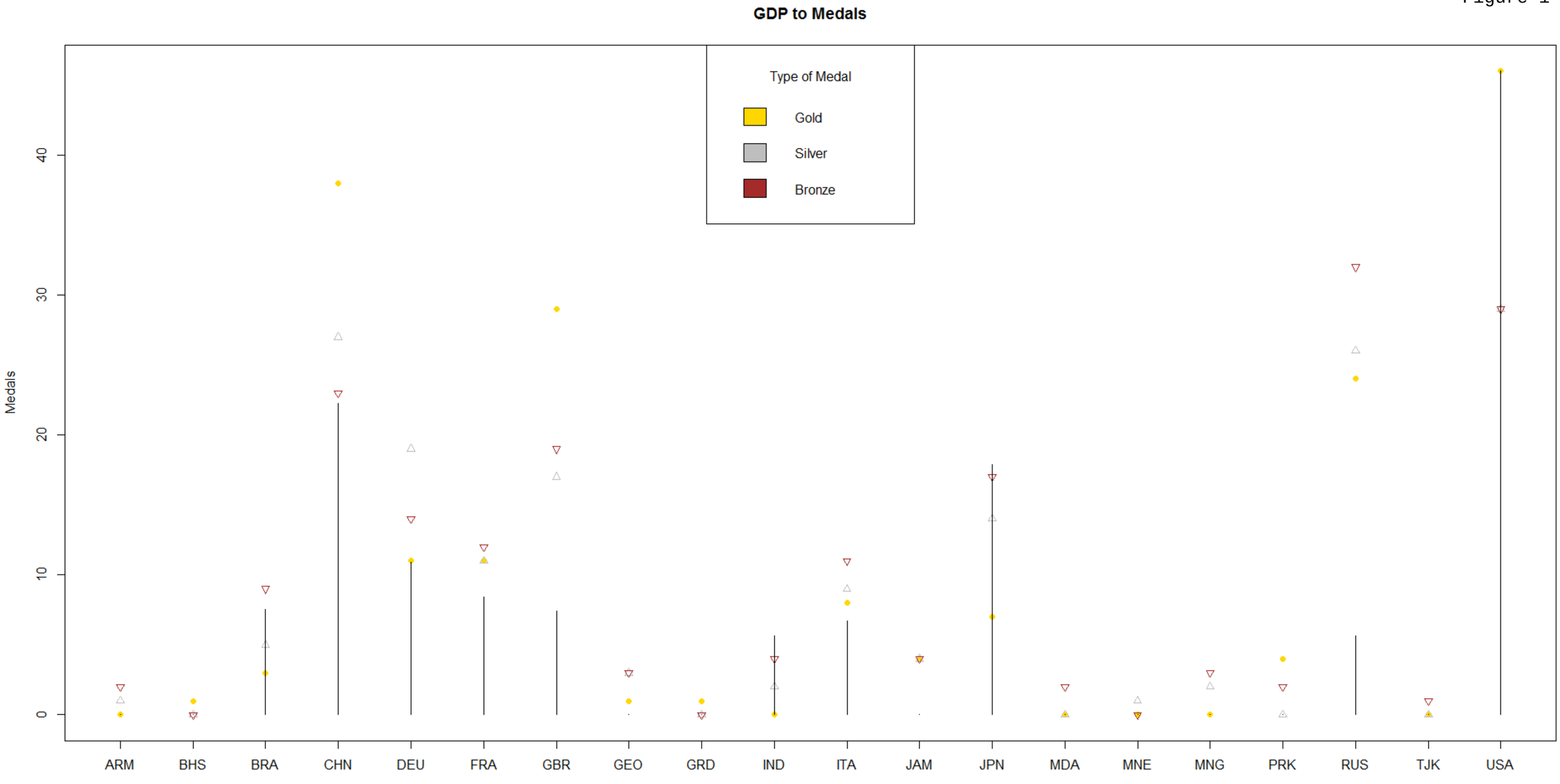
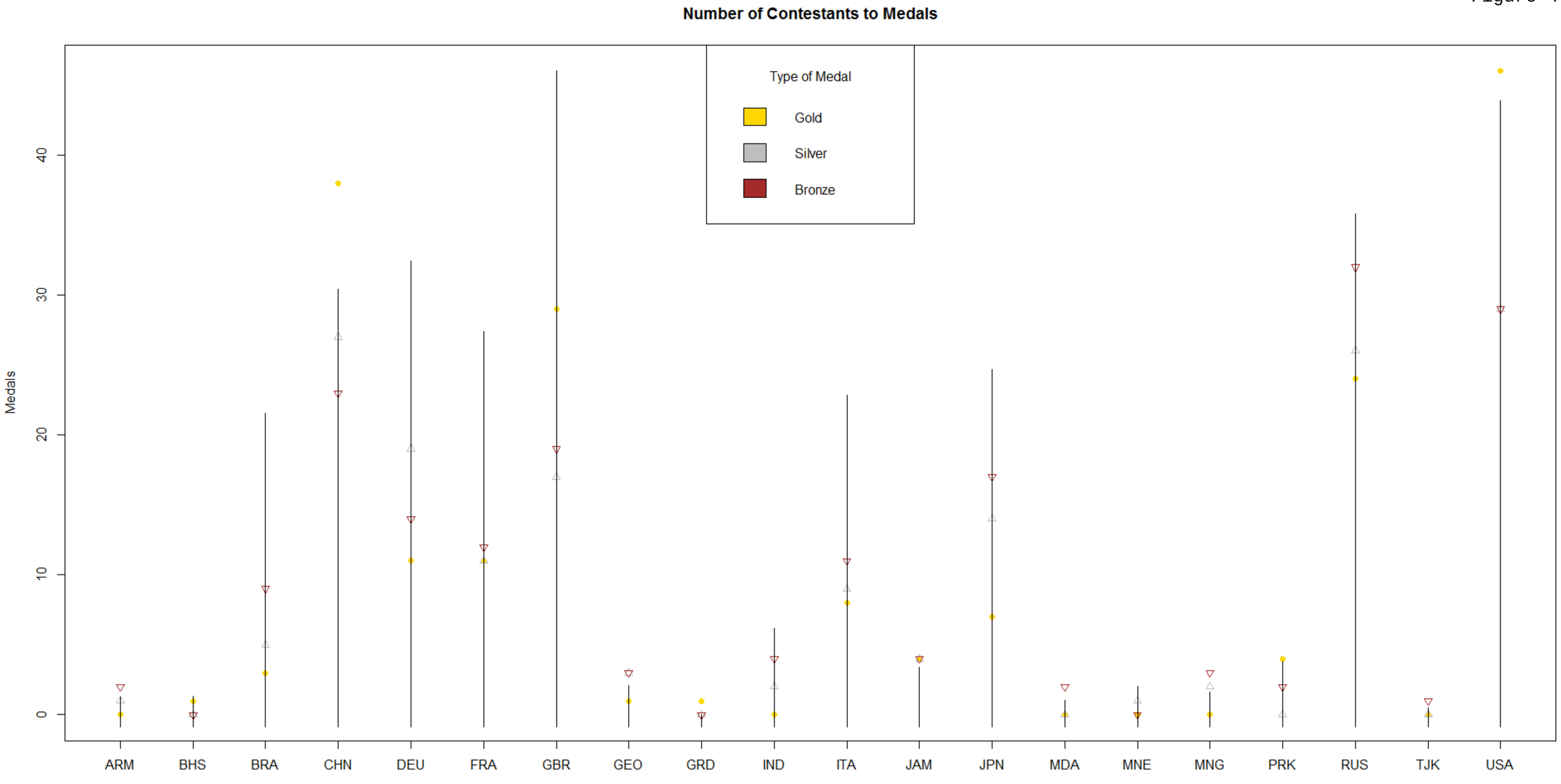

GDP to Medals
Type of Medal
Gold
Silver
Bronze
Medals
0
10
20
30
40
ARM
BHS
BRA
CHN
DEU
FRA
GBR
GEO
GRD
IND
ITA
JAM
JPN
MDA
MNE
MNG
PRK
RUS
TJK
USA

Figure 4


Number of Contestants to Medals
Type of Medal
Gold
Silver
Bronze
Medals
0
10
20
30
40
ARM
BHS
BRA
CHN
DEU
FRA
GBR
GEO
GRD
IND
ITA
JAM
JPN
MDA
MNE
MNG
PRK
RUS
TJK
USA