Spot_plasticity

In this document we will assess the hypotheses tested in the manuscript concerning our experimental populations and the differences in plasticity between native and invasive populations. Our predictions are:

1. **The spot size is obviously correlated to wing size: they both will decrease as temperature increases. Among populations, we do not expect significant diffences in wing size nor spot size within the same temperature.**
2. **With the change in temperature we would expect the Sokol population to have a response more plastic than Sapporo population in wing size (based in Antoine et al. 2018): so wings will be larger for the American population at 16? and smaller at 28?. This pattern may therefore be reproduced for the absolute spot size** since it is highly correlated to wing size.

Our data consists in a data.frame consisting in an individual label, the genetic lineage (population) to which it belongs, the temperature at which it has been raised, its wing size, its spot size and the spot size/wing size ratio. Because wings from the same individual are not independent observations, we estimate the average for each trait:

```
spot.raw <- read.table(paste('/Users/ceferino_vg/Documents/Drosophila_suzukii/',
                             'Spot_evolution/data/results_plasticity.txt',
                             sep=''),
                       header=T, sep=',')

spot.data <- aggregate(Wing ~ IND*TEMP+POP, FUN = mean, data = spot.raw)

spot.data$Spot <- aggregate(Spot ~ IND*TEMP+POP, FUN = mean, data = spot.raw)$Spot

spot.data$Ratio <- aggregate(Ratio ~ IND*TEMP+POP, FUN = mean, data = spot.raw)$Ratio

spot.data$TEMP <- factor(spot.data$TEMP)

head(spot.data)
```

```
##   IND TEMP   POP    Wing    Spot      Ratio
## 1 L04   16 PARIS 1986144 92220.0 0.04644350
## 2 L06   16 PARIS 1946261 84597.0 0.04346038
## 3 L07   16 PARIS 2026222 78718.5 0.03884989
## 4 L09   16 PARIS 1891875 74713.0 0.03949119
## 5 L11   16 PARIS 2006696 94066.5 0.04691930
## 6 L13   16 PARIS 1942214 90485.5 0.04658440
```

For future analyses, we import the libraries we are going to use and we write the function to estimate Cohen's d as an effect size measure:

```
library(ggplot2)
library(ggsignif)
library(wesanderson)

cohens.d <- function(pop1, pop2) {
  d <- (mean(pop1)-mean(pop2))/
    sqrt(((length(pop1)-1)*var(pop1)+(length(pop2)-1)*var(pop2))/(length(pop1)+length(pop2)-2))

  return(d)
}
```

We take different indexes to study the different phenotypes within and among temperatures and populations.

```r
Paris_idx <- which(spot.data$POP=='PARIS')
Sappo_idx <- which(spot.data$POP=='SAPPORO')
Sokol_idx <- which(spot.data$POP=='SOKOL')
temp16_idx <- which(spot.data$TEMP==16)
temp22_idx <- which(spot.data$TEMP==22)
temp28_idx <- which(spot.data$TEMP==28)
```

So we start by studying **1) the relationship between spot size, wing size and temperature**:

```r
summary(aov(Spot ~ Wing+TEMP, data = spot.data))
```

```
##              Df    Sum Sq   Mean Sq F value Pr(>F)
## Wing          1 3.608e+10 3.608e+10 345.667 <2e-16 ***
## TEMP          2 1.041e+09 5.205e+08   4.986 0.0083 **
## Residuals   121 1.263e+10 1.044e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lm(Spot ~ Wing, data = spot.data))$r.squared
```

```
## [1] 0.7252173
```

```r
cohens.d(spot.data$Spot[temp16_idx], spot.data$Spot[temp22_idx])
```

```
## [1] 1.543774
```

```r
cohens.d(spot.data$Spot[temp22_idx], spot.data$Spot[temp28_idx])
```

```
## [1] 2.107001
```

```r
summary(lm(Spot ~ TEMP, data = spot.data))$r.squared
```

```
## [1] 0.6693673
```

As expected, spot size is highly associated to the wing size and temperature. There's a high and obvious correlation between spot size, wing size and temperature (look at Fig 1), which does not make advisable to add both factors in a single model due to collinearity problems.

Nonetheless, it is also interesting to look at the highly significant interaction between temperature and wing size, which reveals that the relationship between the spot and wing sizes depend on the temperature. We can re-test that relationship just for the interaction and visualise the relationship between spot and wing sizes for different temperatures:

```r
summary(aov(Spot ~ Wing:TEMP, data = spot.data))
```

```
##              Df    Sum Sq   Mean Sq F value Pr(>F)
## Wing:TEMP     3 3.703e+10 1.234e+10   117.4 <2e-16 ***
## Residuals   121 1.273e+10 1.052e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lm(Spot ~ Wing:TEMP, data = spot.data))$r.squared
```

```
## [1] 0.7442284
```

```r
summary(lm(Spot ~ Wing, data = spot.data))
```

```
##
## Call:
```

```
## lm(formula = Spot ~ Wing, data = spot.data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -27761  -6461    638   5527  31697
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.362e+03  4.310e+03  -1.476    0.142
## Wing         4.954e-02  2.749e-03  18.017   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10540 on 123 degrees of freedom
## Multiple R-squared:  0.7252, Adjusted R-squared:  0.723
## F-statistic: 324.6 on 1 and 123 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Spot ~ Wing, data = spot.data[temp16_idx,]))$r.squared
```

```
## [1] 0.4976915
```

```r
summary(lm(Spot ~ Wing, data = spot.data[temp22_idx,]))$r.squared
```

```
## [1] 0.3019704
```

```r
summary(lm(Spot ~ Wing, data = spot.data[temp28_idx,]))$r.squared
```

```
## [1] 0.030135
```

```r
s16 <- summary(lm(Spot ~ Wing, data = spot.data[temp16_idx,]))$coefficients
s22 <- summary(lm(Spot ~ Wing, data = spot.data[temp22_idx,]))$coefficients
s28 <- summary(lm(Spot ~ Wing, data = spot.data[temp28_idx,]))$coefficients

db <- (s22[2,1]-s16[2,1])
sd <- sqrt(s22[2,2]^2+s16[2,2]^2)
df <- (lm(Spot ~ Wing, data = spot.data[temp16_idx,])$df.residual+lm(Spot ~ Wing,
                                      data = spot.data[temp22_idx,])$df.residual)
td <- db/sd
td
```

```
## [1] -1.472093
```

```r
df
```

```
## [1] 73
```

```r
2*pt(-abs(td), df)
```

```
## [1] 0.1452941
```

```r
db <- (s28[2,1]-s22[2,1])
sd <- sqrt(s28[2,2]^2+s22[2,2]^2)
df <- (lm(Spot ~ Wing, data = spot.data[temp22_idx,])$df.residual+lm(Spot ~ Wing,
                                      data = spot.data[temp28_idx,])$df.residual)
td <- db/sd
td
```

```
## [1] -0.4172546
```

```r
df
```

```
## [1] 88
```

```r
2*pt(-abs(td), df)
```

```
## [1] 0.6775084
```

```r
db <- (s28[2,1]-s16[2,1])
sd <- sqrt(s28[2,2]^2+s16[2,2]^2)
df <- (lm(Spot ~ Wing, data = spot.data[temp16_idx,])$df.residual+lm(Spot ~ Wing,
                                      data = spot.data[temp28_idx,])$df.residual)
td <- db/sd
td
```

```
## [1] -1.205815
```

```r
df
```

```
## [1] 77
```

```r
2*pt(-abs(td), df)
```

```
## [1] 0.2315809
```

```r
summary(lm(Spot ~ Wing, data = spot.data[temp16_idx,]))
```

```
##
## Call:
## lm(formula = Spot ~ Wing, data = spot.data[temp16_idx, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21227  -5490   1082   5325  19164
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.767e+04  2.867e+04  -2.360   0.0247 *
## Wing         7.869e-02  1.420e-02   5.542 4.54e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9410 on 31 degrees of freedom
## Multiple R-squared:  0.4977, Adjusted R-squared:  0.4815
## F-statistic: 30.72 on 1 and 31 DF,  p-value: 4.538e-06
```

```r
summary(lm(Spot ~ Wing, data = spot.data[temp22_idx,]))
```

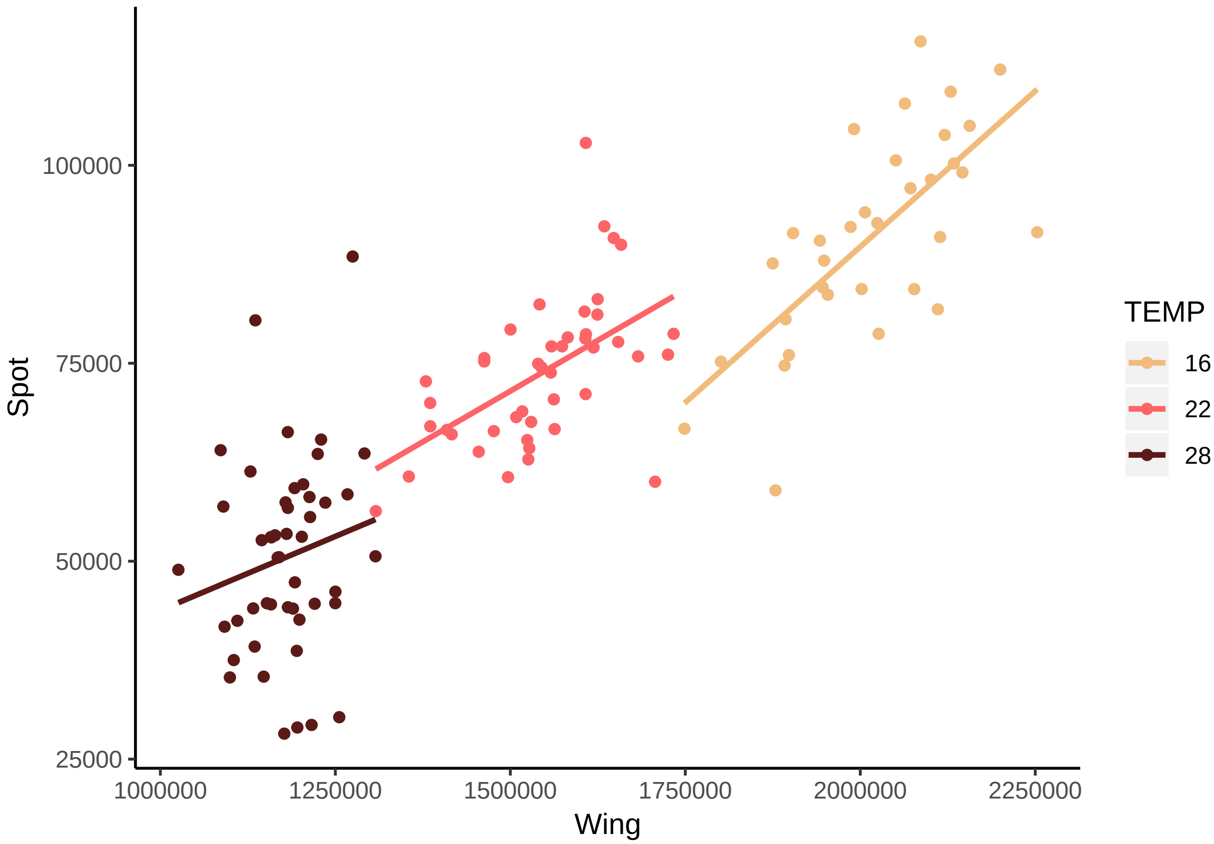
```
##
## Call:
## lm(formula = Spot ~ Wing, data = spot.data[temp22_idx, ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22059.9  -5063.0   -425.2   4429.2  25815.2
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.450e+03  1.863e+04  -0.292 0.771345
```

```
## Wing         5.129e-02  1.203e-02   4.263 0.000112 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7953 on 42 degrees of freedom
## Multiple R-squared:  0.302,  Adjusted R-squared:  0.2854
## F-statistic: 18.17 on 1 and 42 DF,  p-value: 0.0001118
```

```
summary(lm(Spot ~ Wing, data = spot.data[temp28_idx,]))
```

```
##
## Call:
## lm(formula = Spot ~ Wing, data = spot.data[temp28_idx, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -23044  -7089   1076   6505  34428
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.469e+03  3.690e+04   0.175    0.862
## Wing        3.733e-02  3.122e-02   1.196    0.238
##
## Residual standard error: 12300 on 46 degrees of freedom
## Multiple R-squared:  0.03013,    Adjusted R-squared:  0.009051
## F-statistic: 1.429 on 1 and 46 DF,  p-value: 0.238
```

```
ggplot(data = spot.data,
       aes(x = Wing, y = Spot, color = TEMP)) +
  geom_point() +
  geom_smooth(method ='lm', se = FALSE) +
  scale_color_manual(values = wes_palette(n=3, name = "GrandBudapest1")) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
          panel.background = element_blank(), axis.line = element_line(colour = "black"))
```



We can see that the smaller the temperature, the faster the spot size increases with wing size. In addition, we see that the intercept is higher for the populations at 22 and 16 degrees than for the 28 population. This means that we would predict that on average the individual with the smaller wing in the population of 16 or 22 degrees will still have a larger spot than an individual with the largest wing in the population of 28 degrees. For the two first populations this assumption does not hold: on average the smaller individual at 16? will have a smaller spot than the biggest individual at 22?.

To test for **2) differences in plasticity between the native and invasive populations**, we run an ANOVA where the interaction temperature:population will tell us whether there are differences among populations in the effect of temperature on spot size:

```
summary(aov(Spot ~ TEMP*POP, data = spot.data))
```

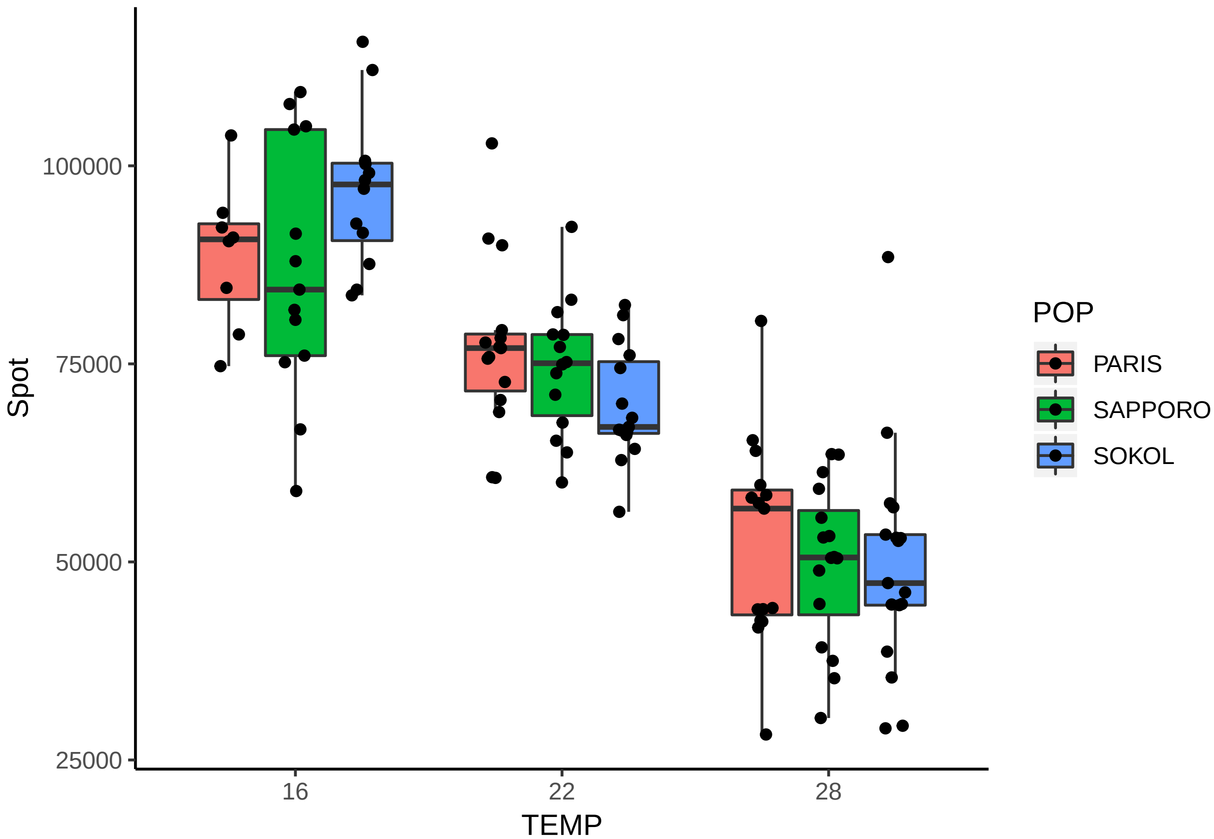
```
##              Df    Sum Sq   Mean Sq F value Pr(>F)
## TEMP          2 3.330e+10 1.665e+10 126.553 <2e-16 ***
## POP           2 1.724e+08 8.620e+07   0.655  0.521
## TEMP:POP      4 1.015e+09 2.537e+08   1.928  0.110
## Residuals   116 1.526e+10 1.316e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(Spot ~ TEMP:POP, data = spot.data))$r.squared
```

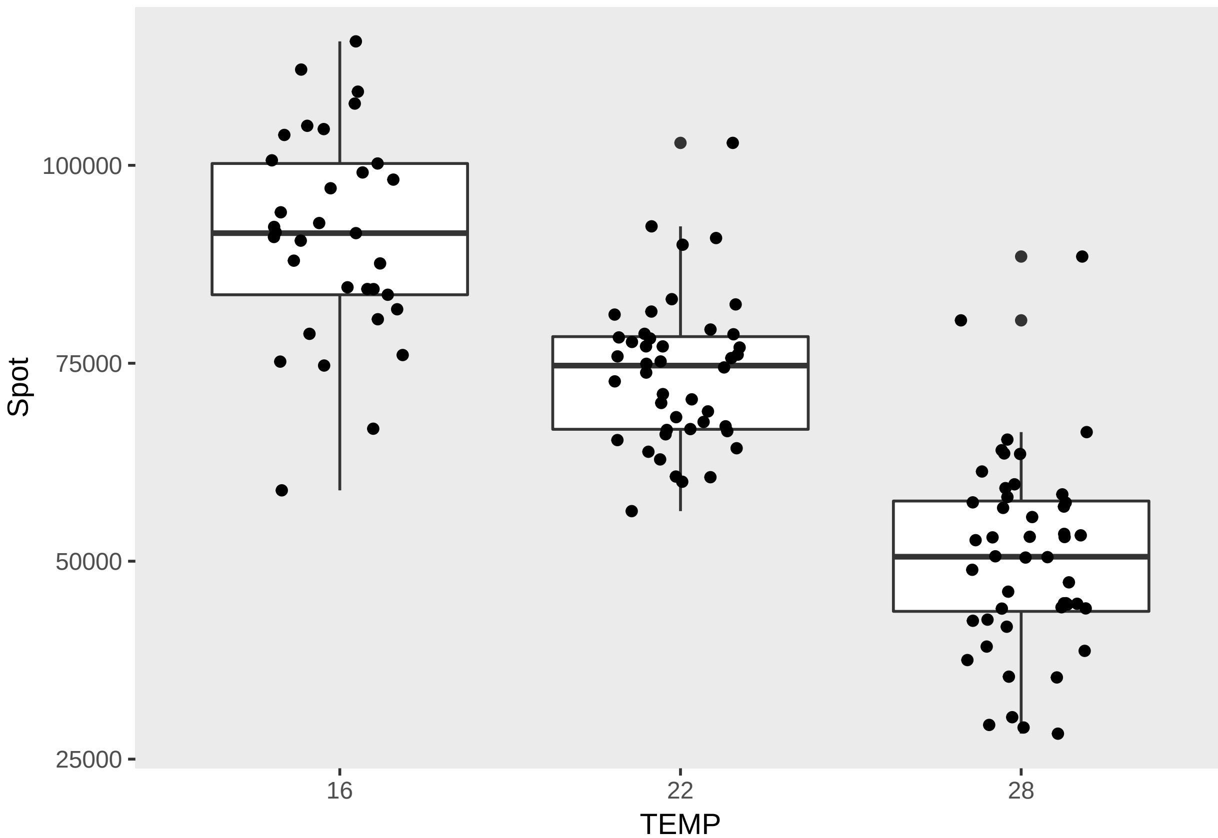
```
## [1] 0.6932242
```

```
ggplot(data = spot.data,
       aes(x = TEMP, y = Spot, fill = POP)) +
    geom_boxplot(outlier.shape = NA) +
    geom_jitter(position = position_jitterdodge(jitter.width = 0.2)) +
```

```
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))
```




```
ggplot(data = spot.data,
       aes(x = TEMP, y = Spot)) +
  geom_boxplot() + geom_jitter(width = 0.2) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
```



So we don't find differences among geographic populations but there are differences in plasticity among them. That doesn't mean, however, that there are differences between invasive and native populations. We look at the R2 for each population independently:

```
summary(lm(Spot ~ TEMP, data = spot.data[Paris_idx,]))$r.squared
```

```
## [1] 0.6375041
```

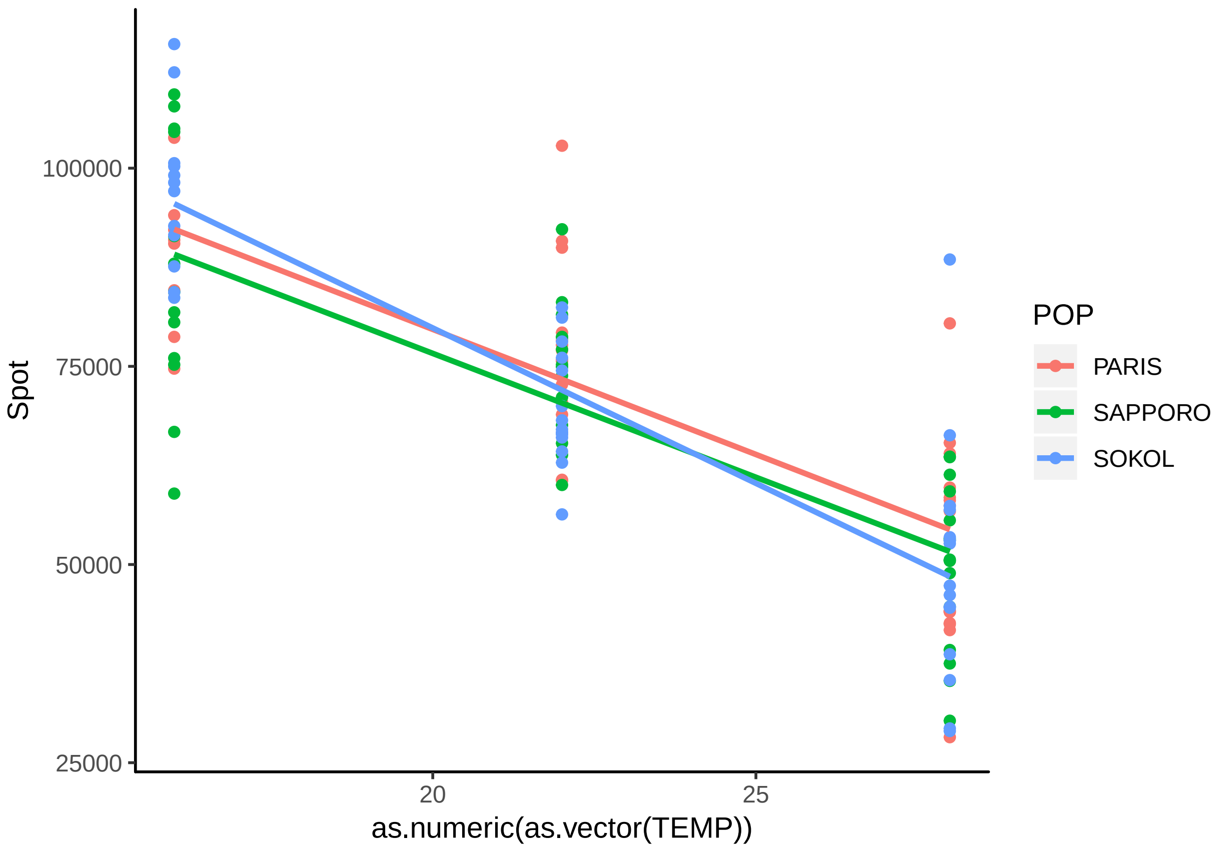
```
summary(lm(Spot ~ TEMP, data = spot.data[Sappo_idx,]))$r.squared
```

```
## [1] 0.6514854
```

```
summary(lm(Spot ~ TEMP, data = spot.data[Sokol_idx,]))$r.squared
```

```
## [1] 0.7594588
```

```
ggplot(data=spot.data,
       aes(x = as.numeric(as.vector(TEMP)), y = Spot, color = POP)) +
  geom_point() +
  geom_smooth(method = 'lm', se = FALSE) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank()) +
  theme(axis.line.x = element_line(color="black", size = 0.5),
        axis.line.y = element_line(color="black", size = 0.5))
```



We find the fastest decrease in spot size with wing size in the Sokol population, while in the Paris population it is the slowest. The Sapporo population stays in between. Therefore, we cannot say that there is a difference between native and invasive populations for the spot size.

We can now repeat the same ANOVA for wing size and the relative spot size to check whether there are differences in plasticity among population for these two traits. We start with the wing size:

```
summary(aov(Wing ~ TEMP*POP, data = spot.data))
```

```
##              Df    Sum Sq   Mean Sq F value   Pr(>F)
## TEMP          2 1.367e+13 6.837e+12 886.278 < 2e-16 ***
## POP           2 1.874e+09 9.371e+08   0.121 0.88572
## TEMP:POP      4 1.347e+11 3.368e+10   4.366 0.00252 **
## Residuals   116 8.948e+11 7.714e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cohens.d(spot.data$Wing[temp16_idx], spot.data$Wing[temp22_idx])
```

```
## [1] 4.355381
```

```
cohens.d(spot.data$Wing[temp22_idx], spot.data$Wing[temp28_idx])
```

```
## [1] 4.498832
```

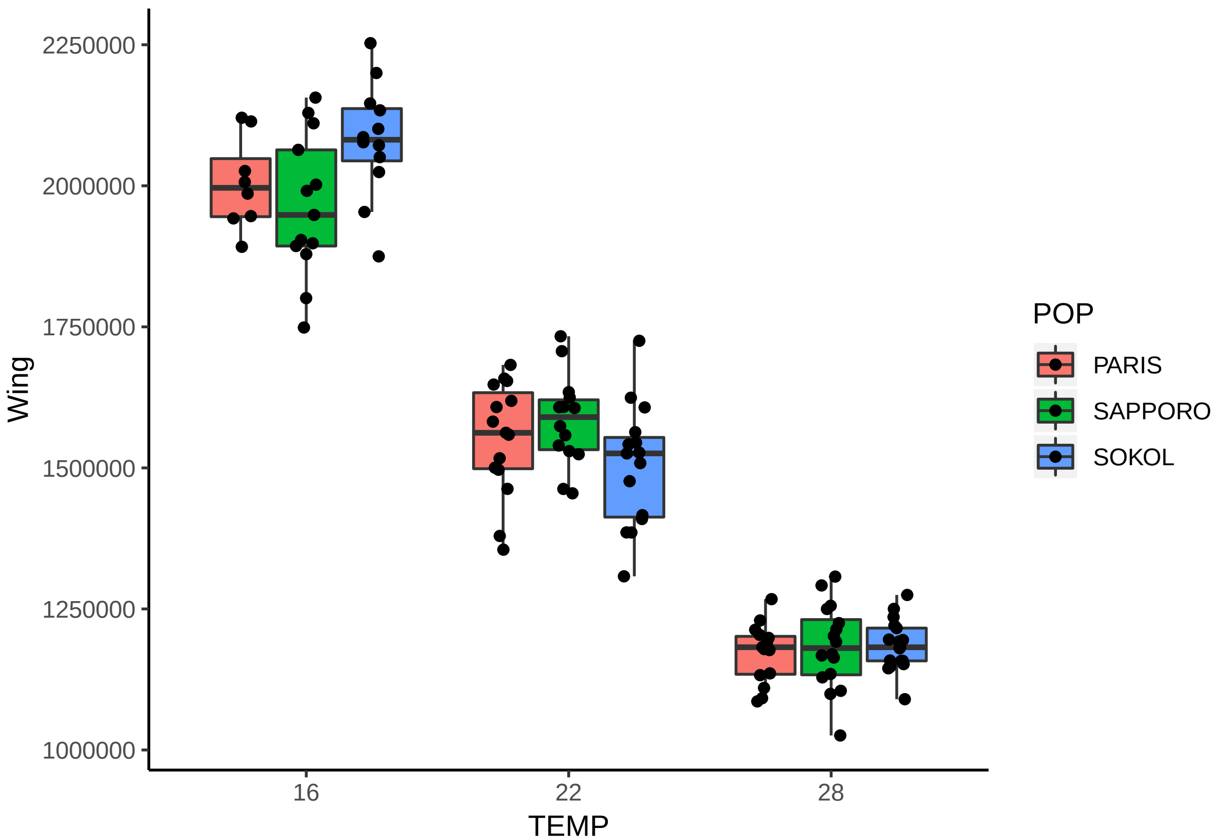
```
summary(lm(Wing ~ TEMP, data = spot.data))$r.squared
```

```
## [1] 0.9298583
```

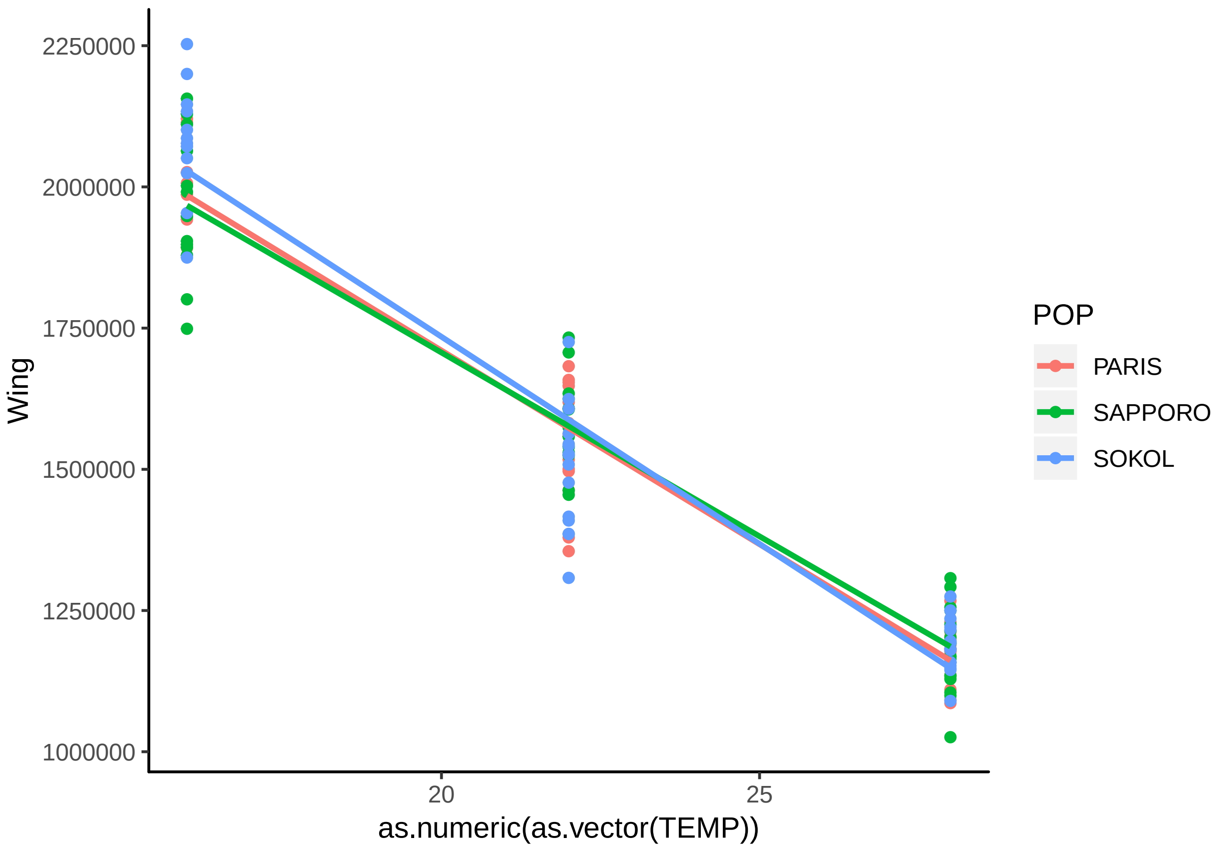
```
summary(lm(Wing ~ TEMP:POP, data = spot.data))$r.squared
```

```
## [1] 0.939148
```

```r
ggplot(data = spot.data,
       aes(x = TEMP, y = Wing, fill = POP)) +
    geom_boxplot(outlier.shape = NA) +
    geom_jitter(position = position_jitterdodge(jitter.width = 0.2)) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
          panel.background = element_blank(), axis.line = element_line(colour = "black"))
```




```r
ggplot(data=spot.data,
       aes(x = as.numeric(as.vector(TEMP)), y = Wing, color = POP)) +
  geom_point() +
  geom_smooth(method = 'lm', se = FALSE) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank()) +
  theme(axis.line.x = element_line(color="black", size = 0.5),
        axis.line.y = element_line(color="black", size = 0.5))
```



As expected, we find an effect of temperature on wing size. And here again we find different plasticities among populations. We look carefully at the R2 for each population to check for differences between native and invasive populations:

```
summary(lm(Wing ~ TEMP, data = spot.data[Paris_idx,]))$r.squared
```

```
## [1] 0.9429256
```

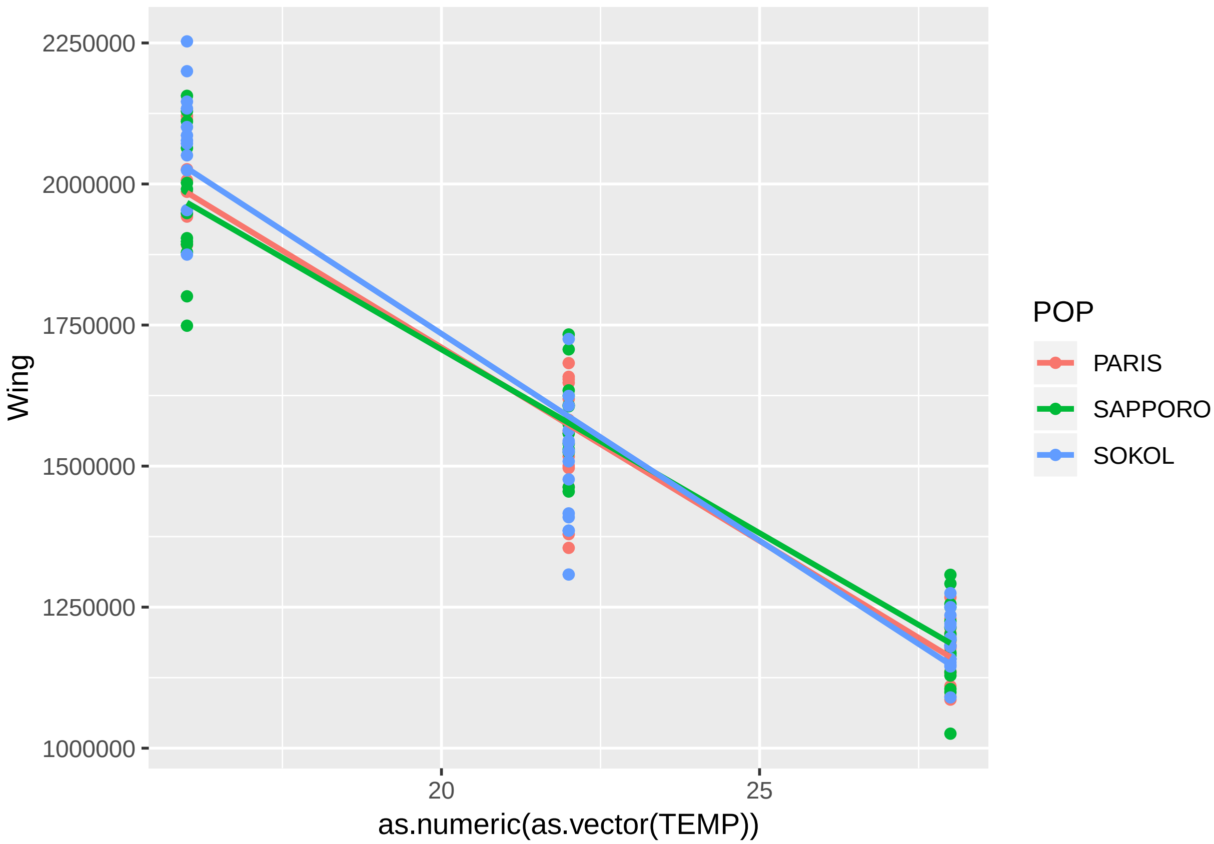
```
summary(lm(Wing ~ TEMP, data = spot.data[Sappo_idx,]))$r.squared
```

```
## [1] 0.9243886
```

```
summary(lm(Wing ~ TEMP, data = spot.data[Sokol_idx,]))$r.squared
```

```
## [1] 0.9477892
```

```
ggplot(data=spot.data,
       aes(x = as.numeric(as.vector(TEMP)), y = Wing, color = POP)) +
    geom_point() +
    geom_smooth(method = 'lm', se = FALSE)
```



This time we did find a difference between the native and invasive populations, although it is minimal. Sokol population decreases its wing size faster with temperature than the other two populations while Sapporo is the slowest and Paris is in between.

We now repeat the ANOVA for the relative spot size on all three different populations for the three different temperatures:

```
summary(aov(Ratio ~ TEMP*POP, data=spot.data))
```

```
##              Df   Sum Sq   Mean Sq F value Pr(>F)
## TEMP          2 0.000555 2.776e-04   4.907  0.009 **
## POP           2 0.000107 5.345e-05   0.945  0.392
## TEMP:POP      4 0.000115 2.875e-05   0.508  0.730
## Residuals   116 0.006562 5.657e-05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cohens.d(spot.data$Ratio[temp16_idx], spot.data$Ratio[temp22_idx])
```

```
## [1] -0.5489807
```

```
cohens.d(spot.data$Ratio[temp22_idx], spot.data$Ratio[temp28_idx])
```

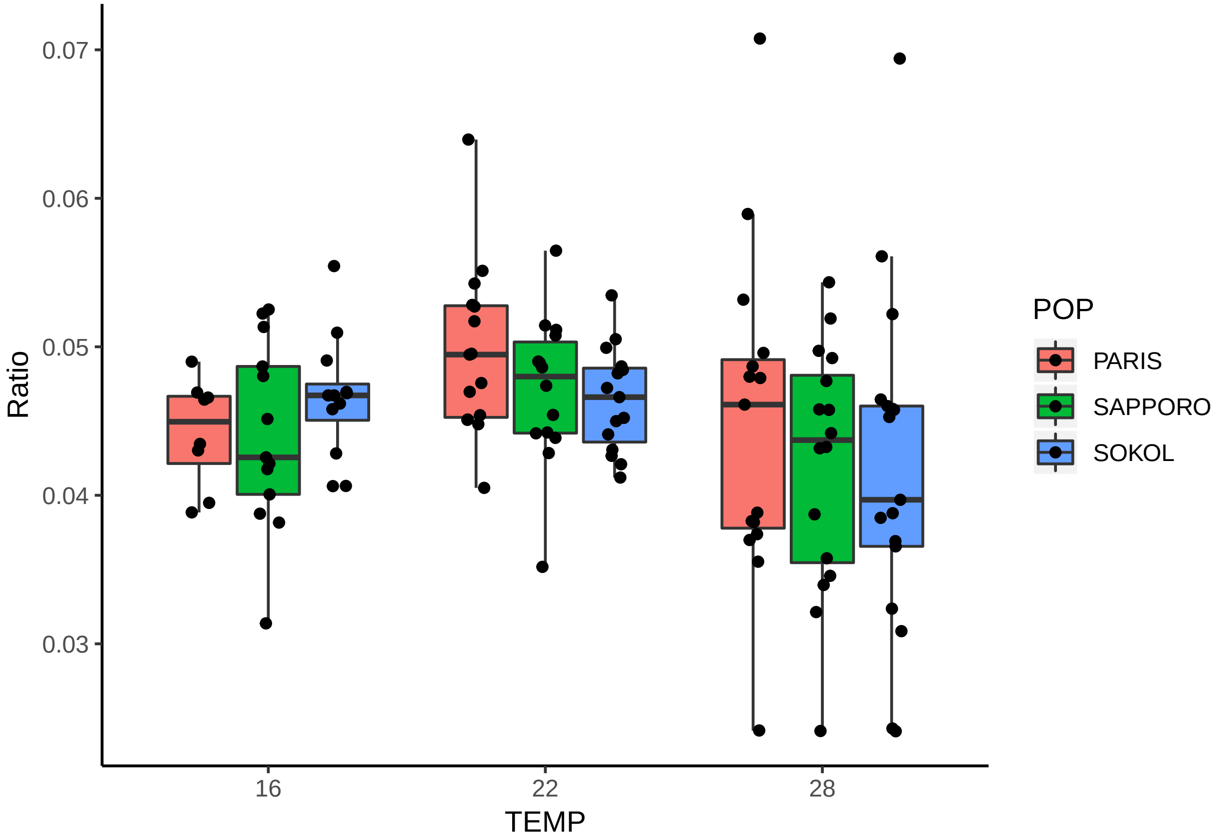
```
## [1] 0.603224
```

```
summary(lm(Ratio ~ TEMP, data=spot.data))$r.squared
```

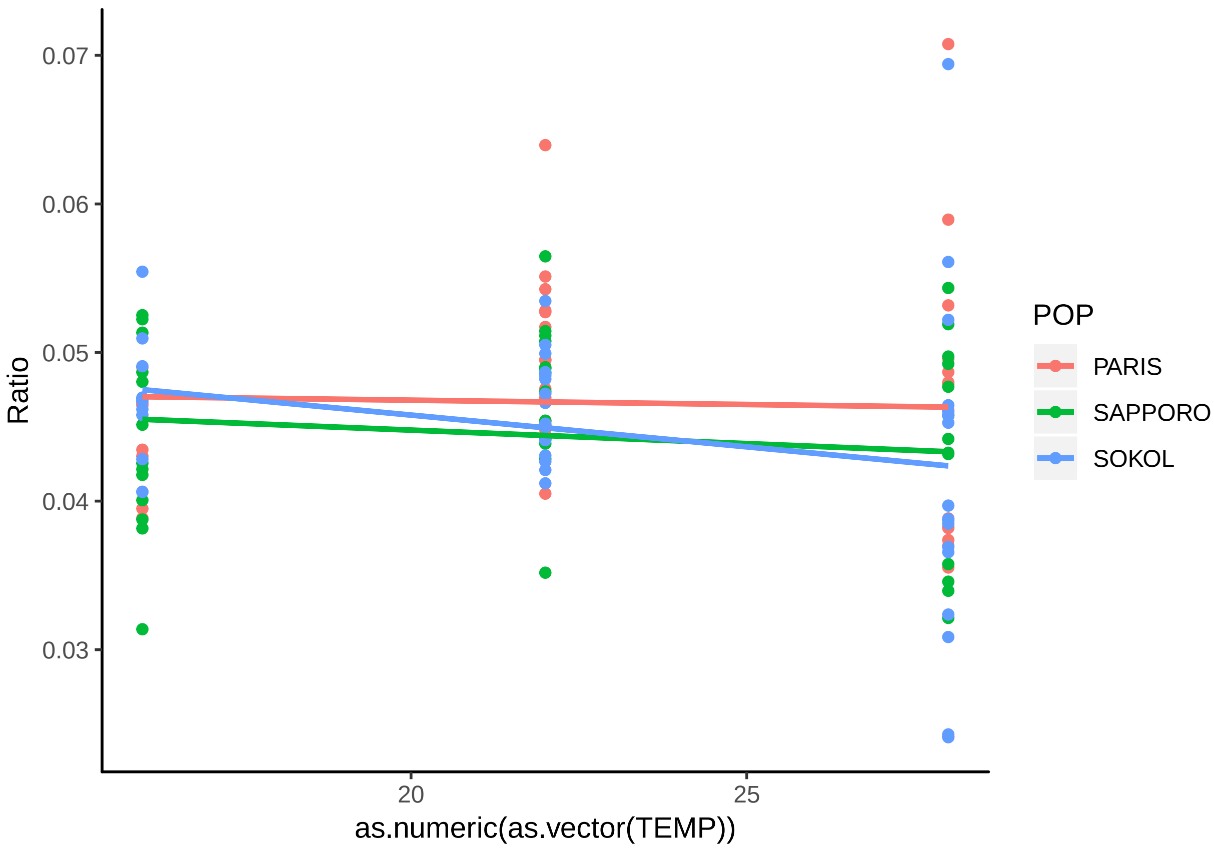
```
## [1] 0.07564782
```

```
ggplot(data = spot.data,
       aes(x = TEMP, y = Ratio, fill = POP)) +
```

```r
    geom_boxplot(outlier.shape = NA) +
    geom_jitter(position = position_jitterdodge(jitter.width = 0.2)) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
          panel.background = element_blank(), axis.line = element_line(colour = "black"))
```




```r
ggplot(data=spot.data,
       aes(x = as.numeric(as.vector(TEMP)), y = Ratio, color = POP)) +
    geom_point() +
    geom_smooth(method = 'lm', se = FALSE) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
          panel.background = element_blank()) +
    theme(axis.line.x = element_line(color="black", size = 0.5),
          axis.line.y = element_line(color="black", size = 0.5))
```



The ANOVA testing for the effect of temperature, population and their interaction shows significant differences for temperature in the case of the relative spot size and for the residual spot size. Everything looks (and shows on the Cohen's d estimates) much more stable than for the absolute spot size and the wing size. We check that, indeed, the 22? temperature has overall larger relative and residual spot size than the other two.:

```
summary(aov(Ratio ~ TEMP*POP, data=spot.data[-temp22_idx,]))
```

```
##             Df   Sum Sq   Mean Sq F value Pr(>F)
## TEMP         1 0.000093 9.278e-05   1.243  0.268
## POP          2 0.000044 2.176e-05   0.292  0.748
## TEMP:POP     2 0.000091 4.562e-05   0.611  0.545
## Residuals   75 0.005596 7.462e-05
```

So, one possibility could be to use from here on the residuals from the linear model *spot size* ~ *wing size* to study the behaviour of the spot size independent to the wing size. We could see this as a similar proxy to the use of the ratio parameter since the latter will also show bigger values when the spot size is anormally bigger in relation to a given wing size.

Indeed, we obtain similar results:

```
spot.ind <- aov(Spot ~ Wing, data=spot.data)

spot.data$Res <- spot.ind$residuals

summary(aov(Res ~ TEMP*POP, data=spot.data))
```

```
##              Df    Sum Sq   Mean Sq F value Pr(>F)
## TEMP          2 9.088e+08 454406251   4.291 0.0159 *
## POP           2 1.863e+08  93173563   0.880 0.4176
```

```
## TEMP:POP      4 2.934e+08  73351700   0.693 0.5985
## Residuals   116 1.228e+10 105890322
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(Res ~ TEMP, data=spot.data))$r.squared
```

```
## [1] 0.06647329
```

```
summary(lm(Res ~ TEMP, data = spot.data[Paris_idx,]))$r.squared
```

```
## [1] 0.146113
```

```
summary(lm(Res ~ TEMP, data = spot.data[Sappo_idx,]))$r.squared
```

```
## [1] 0.07027914
```

```
summary(lm(Res ~ TEMP, data = spot.data[Sokol_idx,]))$r.squared
```

```
## [1] 0.03867912
```

```
cohens.d(spot.data$Res[temp16_idx], spot.data$Res[temp22_idx])
```

```
## [1] -0.7013622
```

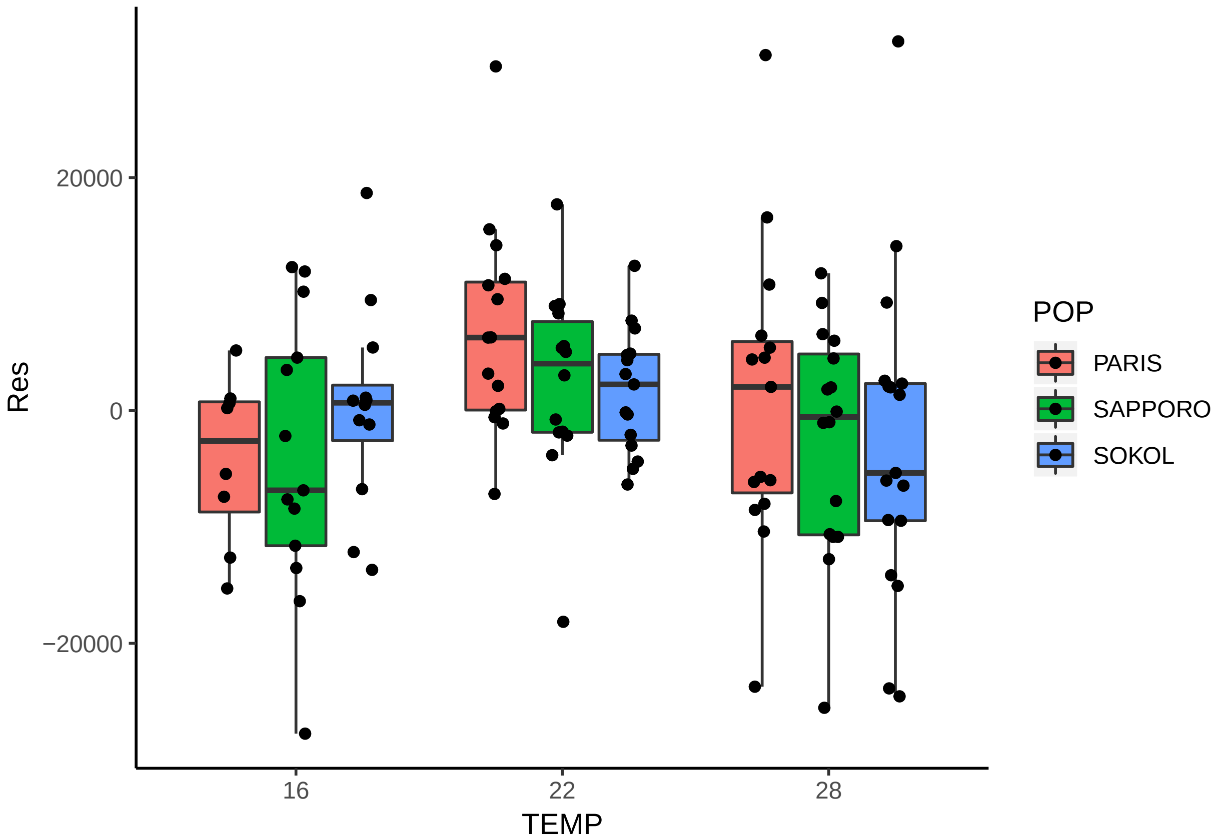
```
cohens.d(spot.data$Res[temp22_idx], spot.data$Res[temp28_idx])
```

```
## [1] 0.5025129
```

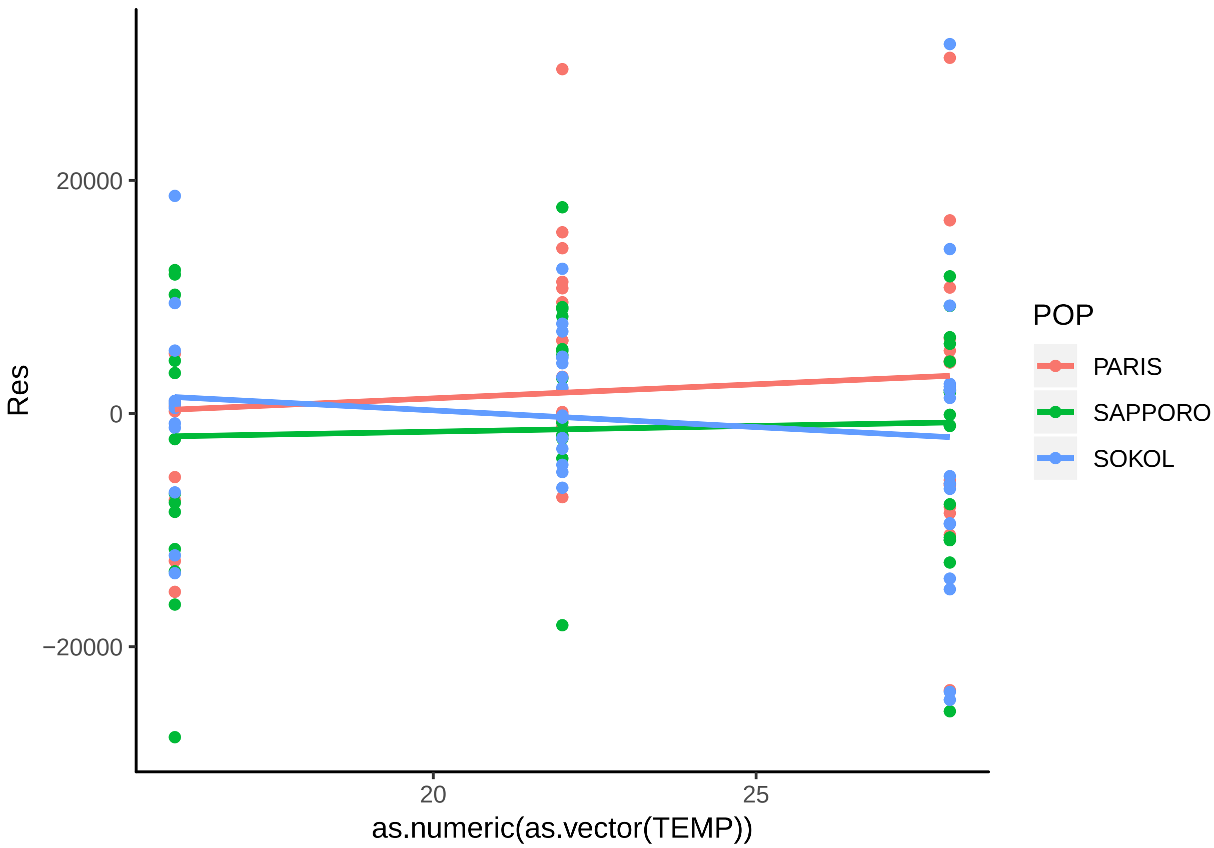
```
summary(aov(Res ~ TEMP*POP, data=spot.data[-temp22_idx,]))
```

```
##             Df    Sum Sq   Mean Sq F value Pr(>F)
## TEMP         1 1.776e+07  17763505   0.135  0.714
## POP          2 6.232e+07  31159118   0.238  0.789
## TEMP:POP     2 2.031e+08 101551858   0.774  0.465
## Residuals   75 9.840e+09 131193833
```

```
ggplot(data = spot.data,
       aes(x = TEMP, y = Res, fill = POP)) +
    geom_boxplot(outlier.shape = NA) +
    geom_jitter(position = position_jitterdodge(jitter.width = 0.2)) +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
          panel.background = element_blank(), axis.line = element_line(colour = "black"))
```




```r
ggplot(data=spot.data,
       aes(x = as.numeric(as.vector(TEMP)), y = Res, color = POP)) +
  geom_point() +
  geom_smooth(method = 'lm', se = FALSE) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank()) +
  theme(axis.line.x = element_line(color="black", size = 0.5),
        axis.line.y = element_line(color="black", size = 0.5))
```




```r
summary(aov(Res ~ TEMP*POP, data=spot.data[-temp22_idx,]))
```

```
##             Df    Sum Sq   Mean Sq F value Pr(>F)
## TEMP         1 1.776e+07  17763505   0.135  0.714
## POP          2 6.232e+07  31159118   0.238  0.789
## TEMP:POP     2 2.031e+08 101551858   0.774  0.465
## Residuals   75 9.840e+09 131193833
```

In summary:

1. The spot size, wing size and temperature are correlated. Temperature decreases both wing and spot sizes, although the spot progressively stops decreasing and therefore the correlation is lost.

2. There isn't evidence about a difference in plasticity between native and invasive populations. This pattern is only followed for the absolute spot size but the differences are minimal.