```
library(tidyverse)
library(knitr)
library(ggplot2)
```

Fruits per plant

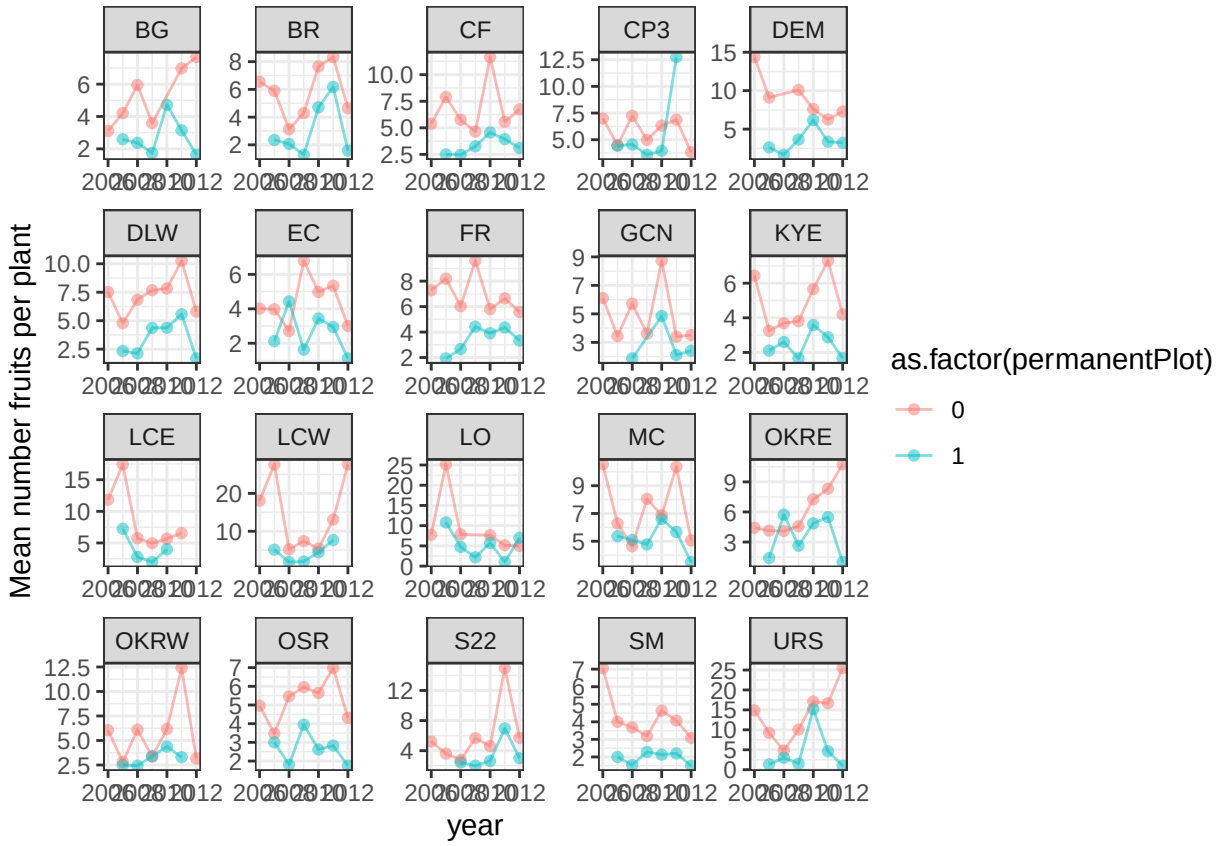
Do plants in permanent plots have more fruits than plants outside of permanent plots? I started thinking about this pattern because I wanted to understand seed rain into plots. Permanent plots may be more likely to be surveyed for all plants, while extra plants are counted ad hoc. This sampling method may bias the 'extra plants' towards plants that are larger, have more fruits, or are more visible. A bias in plant size could matter if seed rain is calculated by using the average number of fruits per plant to estimate plant size in a population-year combination.

I calculated the average number of fruits per plant in permanent plots vs. outside of permanent plots. I did not correct for sampling variation. This first pass suggests that plants outside of permanent plots are, on average, larger than those in permanent plots.

```
countFruitsPerPlantAllPlots <-  readRDS("~/Dropbox/dataLibrary/postProcessingData/countFruitsPerPlantAl
```

```
ggplot(data=countFruitsPerPlantAllPlots %>%
         group_by(site,year,permanentPlot) %>%
         dplyr::summarise(mu=mean(countFruitNumberPerPlant))) +
  geom_point(aes(x=year,y=mu,color=as.factor(permanentPlot)),alpha=.5) +
  geom_line(aes(x=year,y=mu,color=as.factor(permanentPlot)),alpha=.5) +
  facet_wrap(~site,scale='free') +
  ylab("Mean number fruits per plant") +
  theme_bw()
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



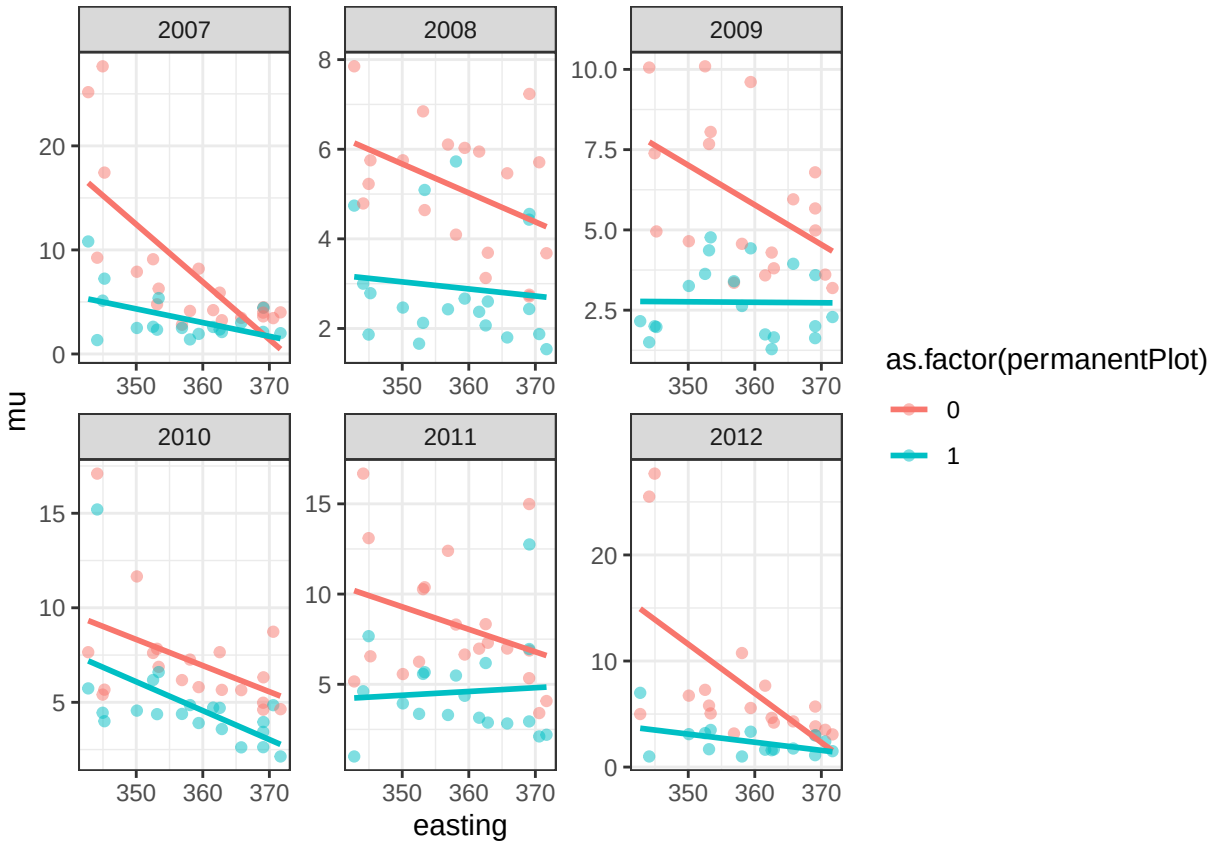
Next, I looked to see whether there is a geographic component to this relationship.

```
position<-read.csv(file="~/Dropbox/projects/clarkiaScripts/data/reshapeData/siteAbiotic.csv",
                   header=TRUE) %>%
  dplyr::select(site,easting) %>%
  dplyr::mutate(easting=easting/1000)

countFruitsPerPlantAllPlotsSummary<-countFruitsPerPlantAllPlots %>%
  dplyr::group_by(site,year,permanentPlot) %>%
  dplyr::summarise(mu = mean(countFruitNumberPerPlant),
                   sd = sd(countFruitNumberPerPlant),
                   n = n(),
                   se = sd/n,
                   p = 1/se) %>%
  dplyr::left_join(position,by="site")
```

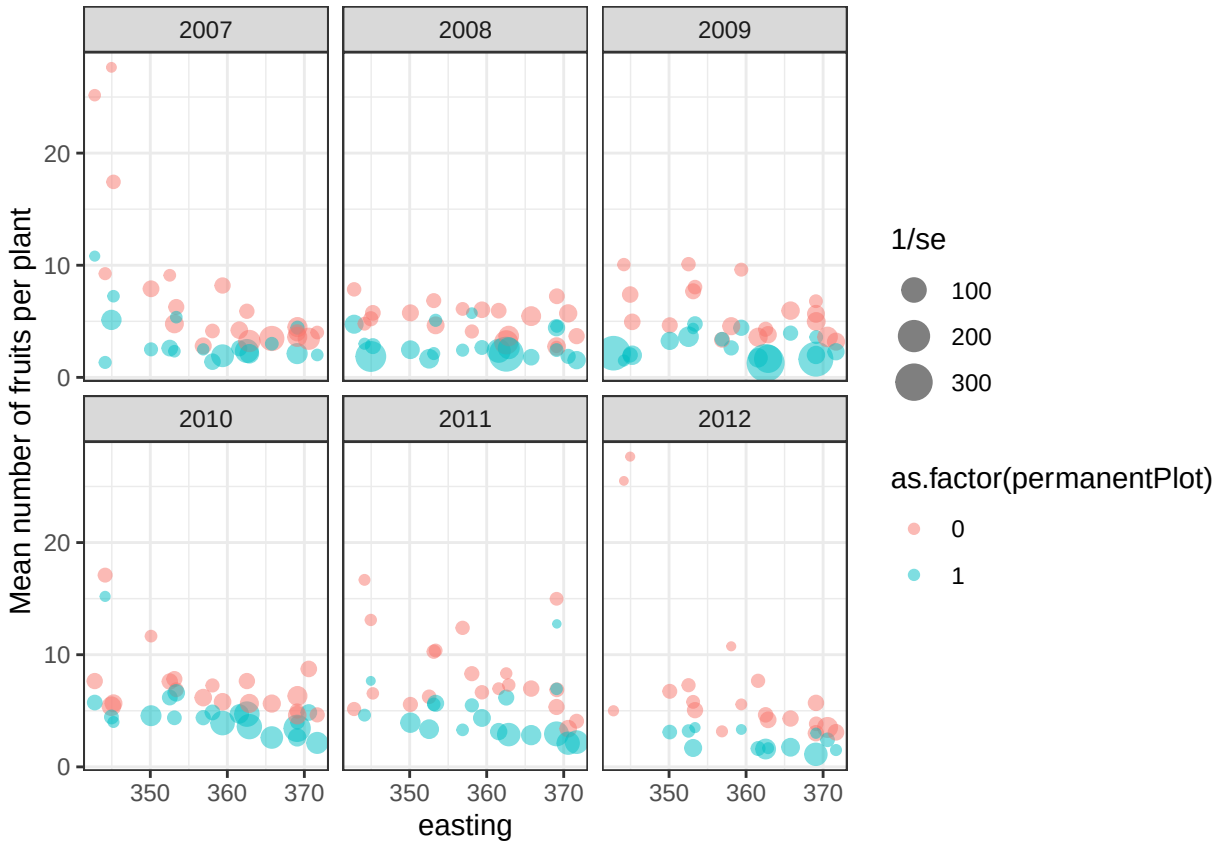
I plotted the geographic pattern to estimates from plants in (blue) and outside (pink) of permanent plots.

```
ggplot(data=countFruitsPerPlantAllPlotsSummary %>% dplyr::filter(year>2006)) +
  geom_point(aes(x=easting,y=mu,color=as.factor(permanentPlot)),alpha=.5) +
  geom_smooth(aes(x=easting,y=mu,color=as.factor(permanentPlot)),method='lm',se=FALSE) +
  facet_wrap(~year,scale='free') +
  theme_bw()
```



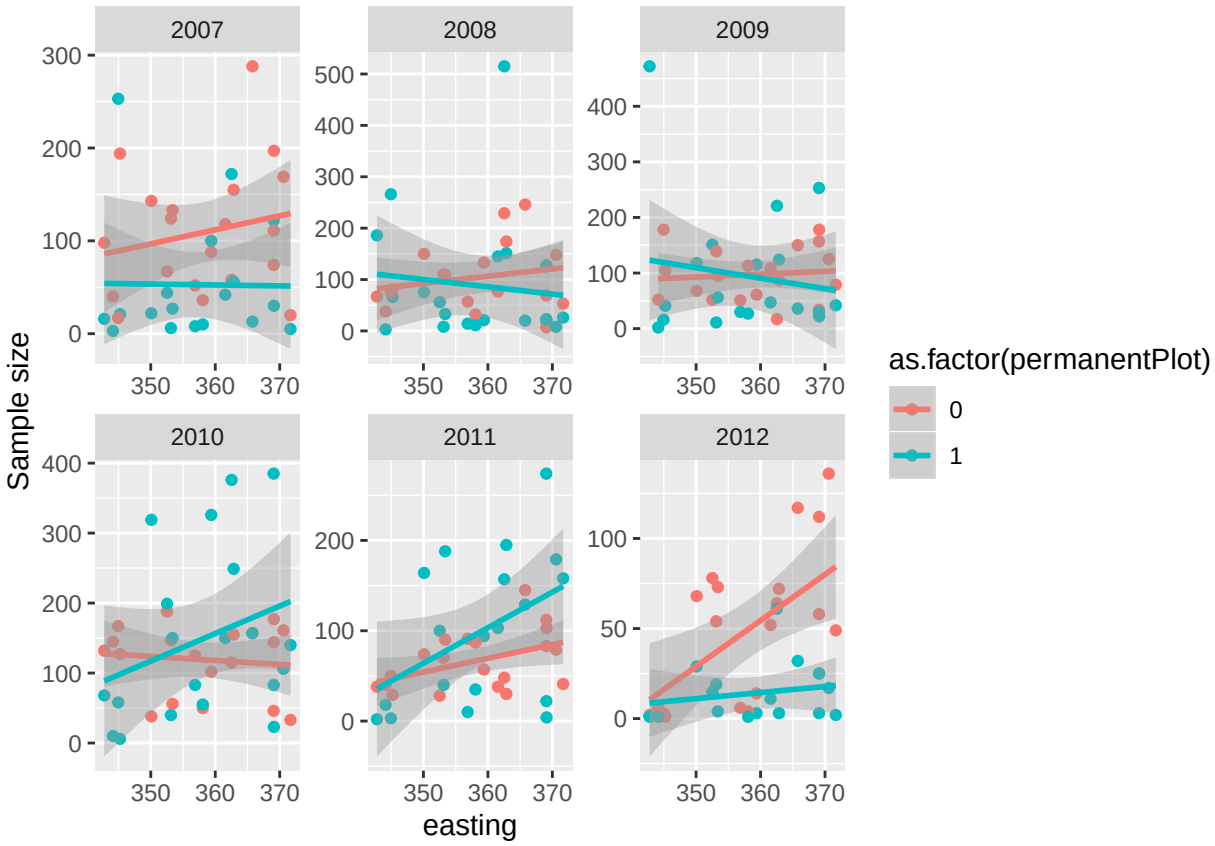
I plotted the mean number of fruits per plant vs. easting, with point size corresonding to 1/standard error. Small points have a higher SE; large points have a smaller SE.

```
ggplot(countFruitsPerPlantAllPlotsSummary %>% dplyr::filter(year>2006)) +
  geom_point(aes(x=easting,y=mu,size=1/se,color=as.factor(permanentPlot)),alpha=.5) +
  facet_wrap(~year) +
  ylab("Mean number of fruits per plant") +
  theme_bw()
```



I looked at how whether the sample size in each category changed geographically.

```
ggplot(countFruitsPerPlantAllPlotsSummary %>% dplyr::filter(year>2006)) +
  geom_point(aes(x=easting,y=n,color=as.factor(permanentPlot))) +
  geom_smooth(aes(x=easting,y=n,color=as.factor(permanentPlot)),method='lm') +
  facet_wrap(~year,scales='free') +
  ylab("Sample size")
```


2007
2008
2009
2010
2011
2012
Sample size
easting
as.factor(permanentPlot)
0
1