`autoplot` : ready made plots with ggplot2

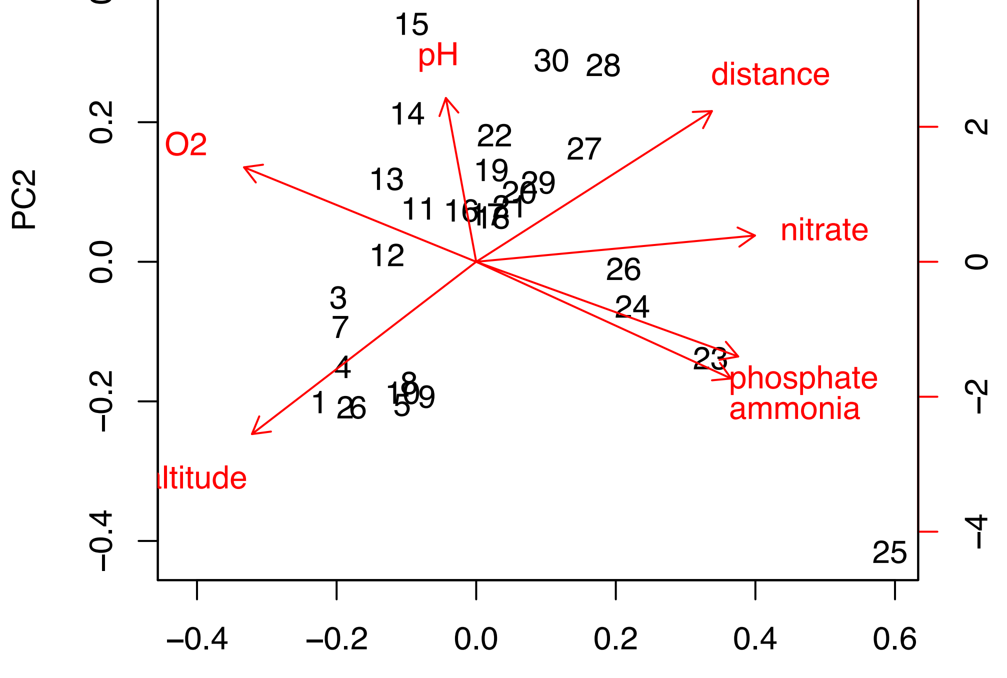
Deuxièmes rencontres R

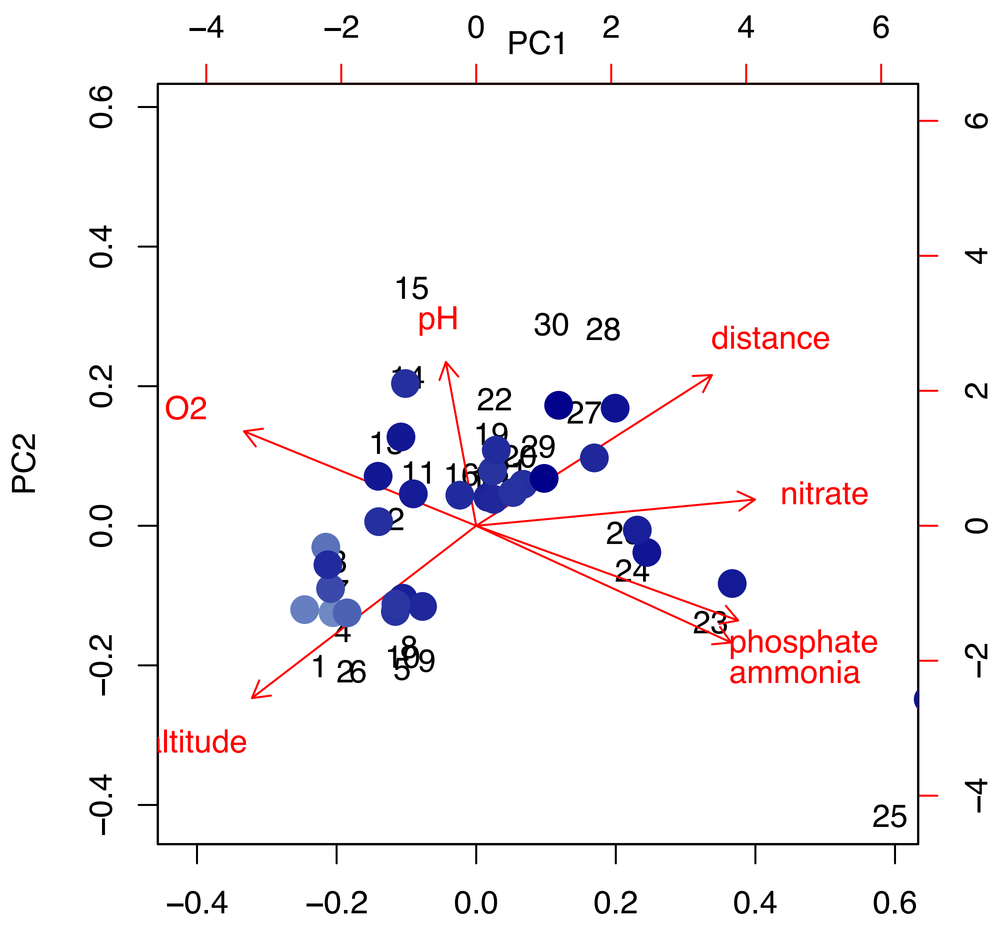
Lyon, 27-28 Juillet 2013

Jean-Olivier Irisson

The problem

```
head(env, 1)
obj <- prcomp(env, scale.=T)

class(obj)
[1] "prcomp"

summary(obj)
biplot(obj)

blues <- colorRamp(
  c("lightblue", "darkblue"))

cols <- rgb(blues(allEnv$hard /
  max(allEnv$hard)), max=255)

points(obj $x[,1], obj $x[,2],
  col=cols, pch=16, cex=2)
```

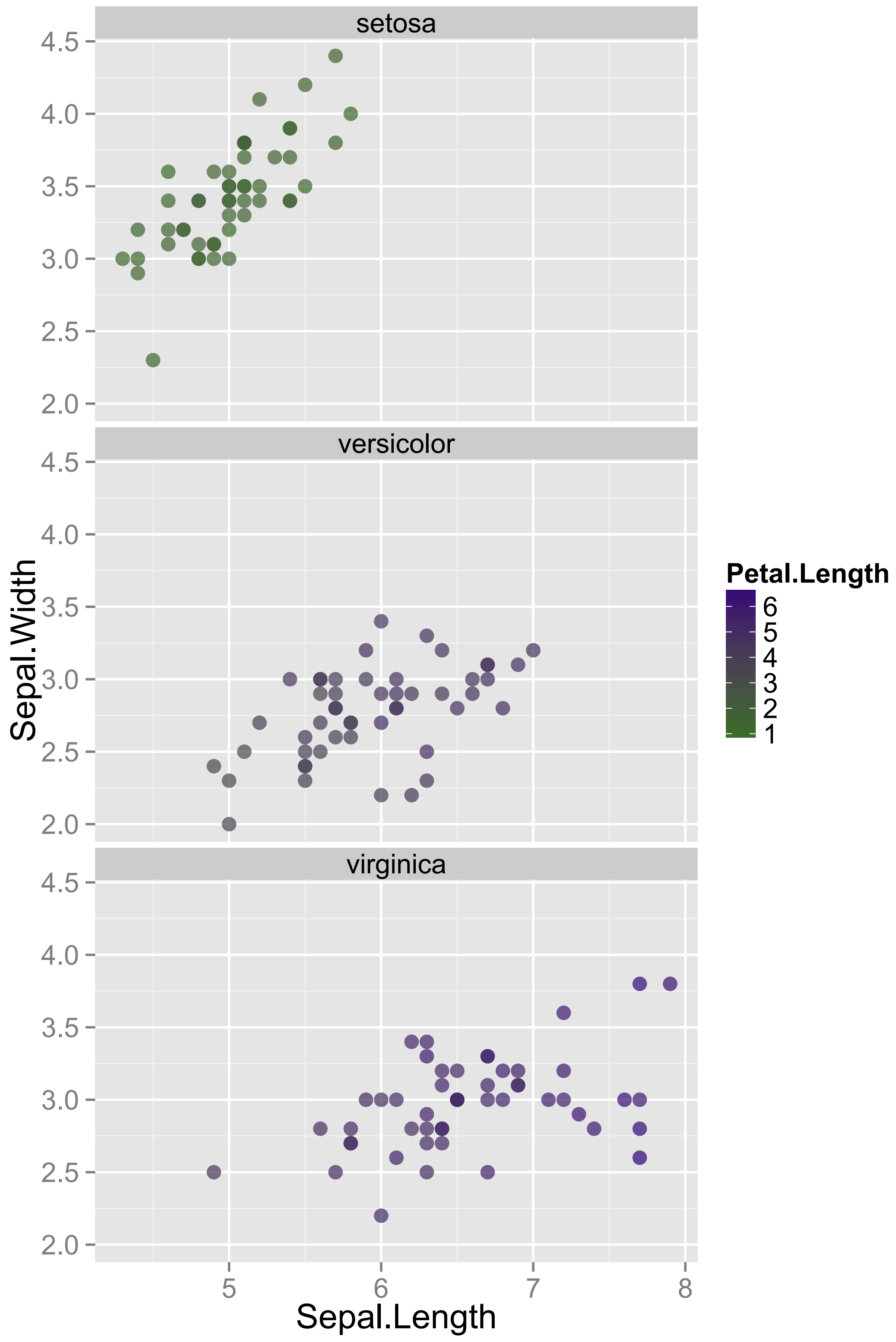




Why ggplot?

- Visually **pleasing**
- **Automatic** everything
- **Customizable** after the fact

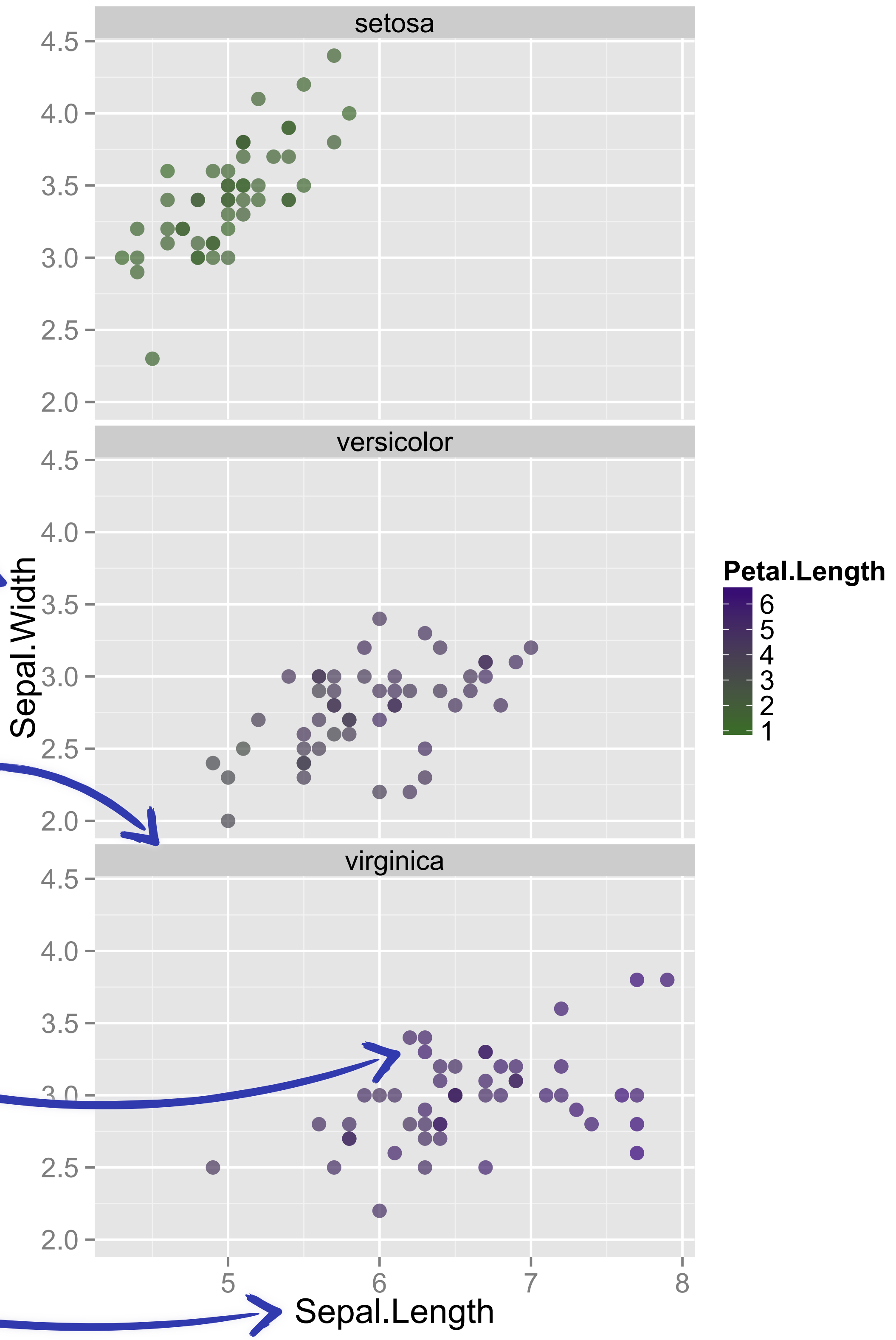
```
p <- ggplot(iris) +
  geom_point(aes(
    x=Sepal.Length,
    y=Sepal.Width,
    colour=Petal.Length),
  alpha=0.7, size=3)

p +
  facet_wrap(~Species, ncol=1)+
  scale_colour_continuous(
    low="#3C6B2A", high="#350079")
```



How does ggplot2 work?

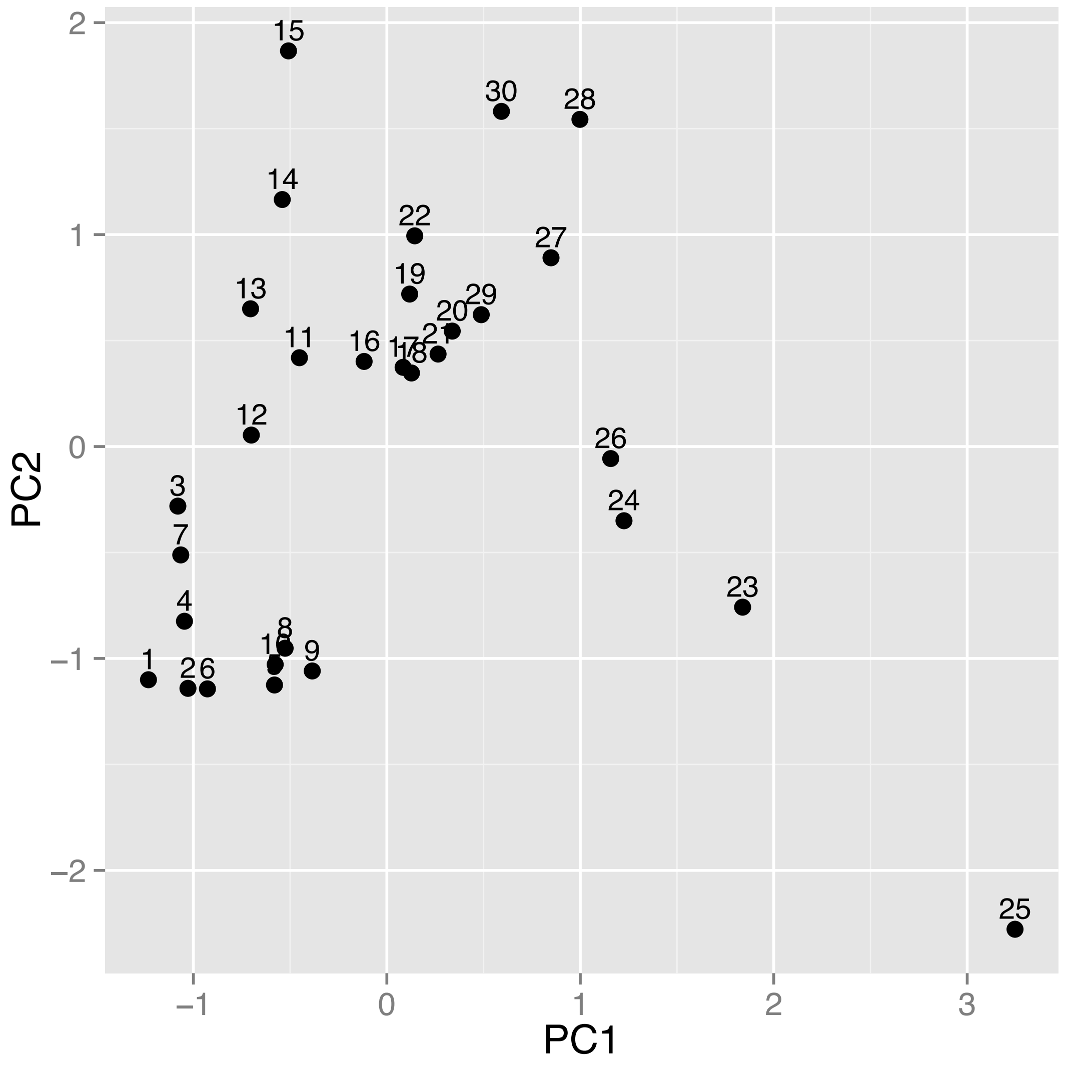
- **grammar** of graphics
- **mapping** data ↔ aesthetics

Sepal. length	Sepal. width	Petal. length	Peta. width	Species
6.0	3.0	4.8	1.8	virginica
6.4	2.9	4.3	1.3	versicol
4.6	3.6	1.0	0.2	setosa
4.3	3.0	1.1	0.1	setosa



Use ggplot2 for multivariate analyses output

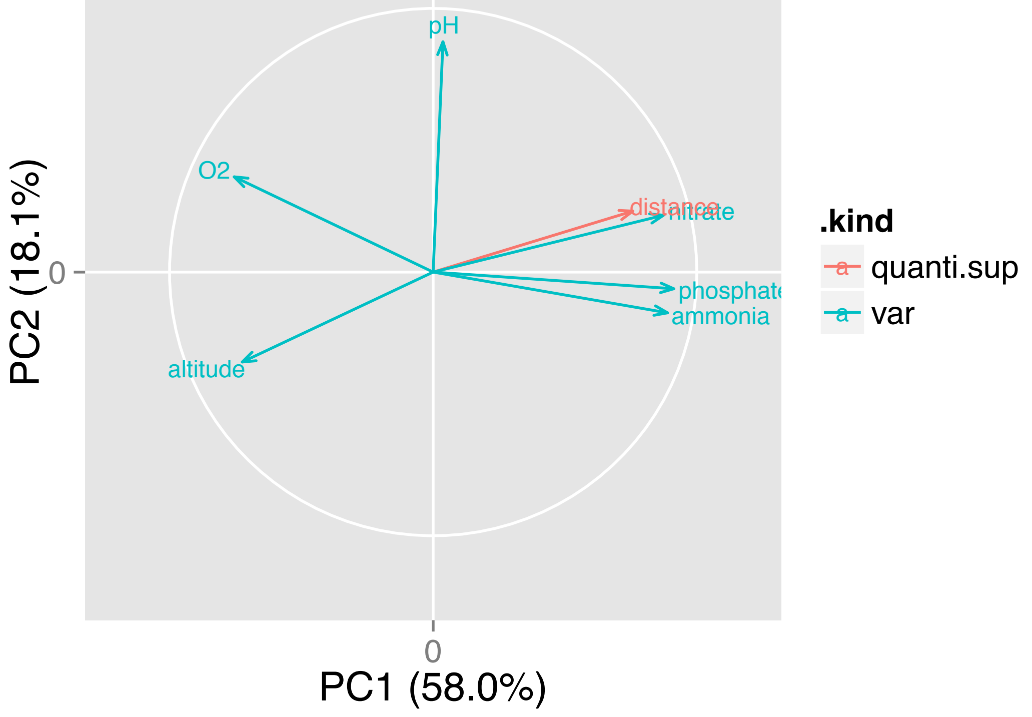
```
obj <- prcomp(env, scale.=T)

# prepare a data.frame
scores <- obj$x[,1:2]
lambda <- obj$sdev[1:2]
d <- data.frame(id=1:nrow(d),
  t(t(scores)/lambda))

# plot it
ggplot(d, aes(x=PC1, y=PC2)) +
  geom_point() +
  geom_text(aes(label=id),
    size=3, vjust=-0.5)

# or
autoplot(obj, type=”obs”)
```

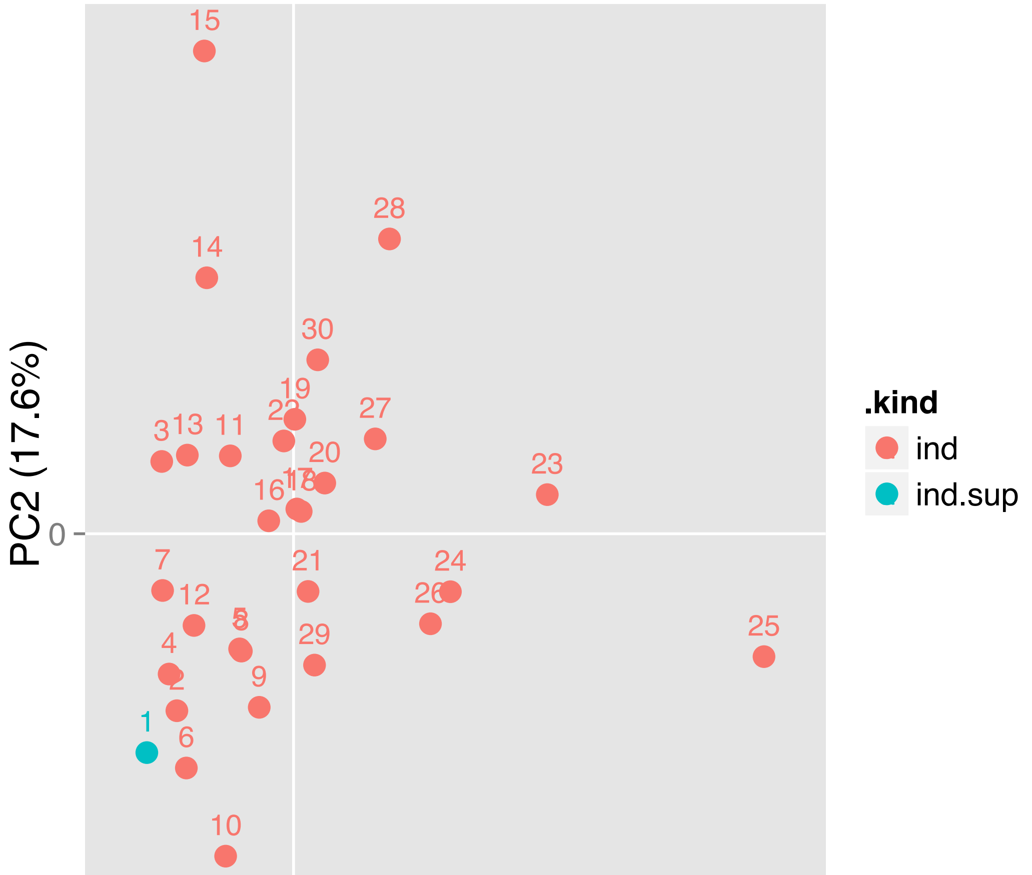


Automatic

- computation of percentage of variance explained
- better scales
- automatic colours

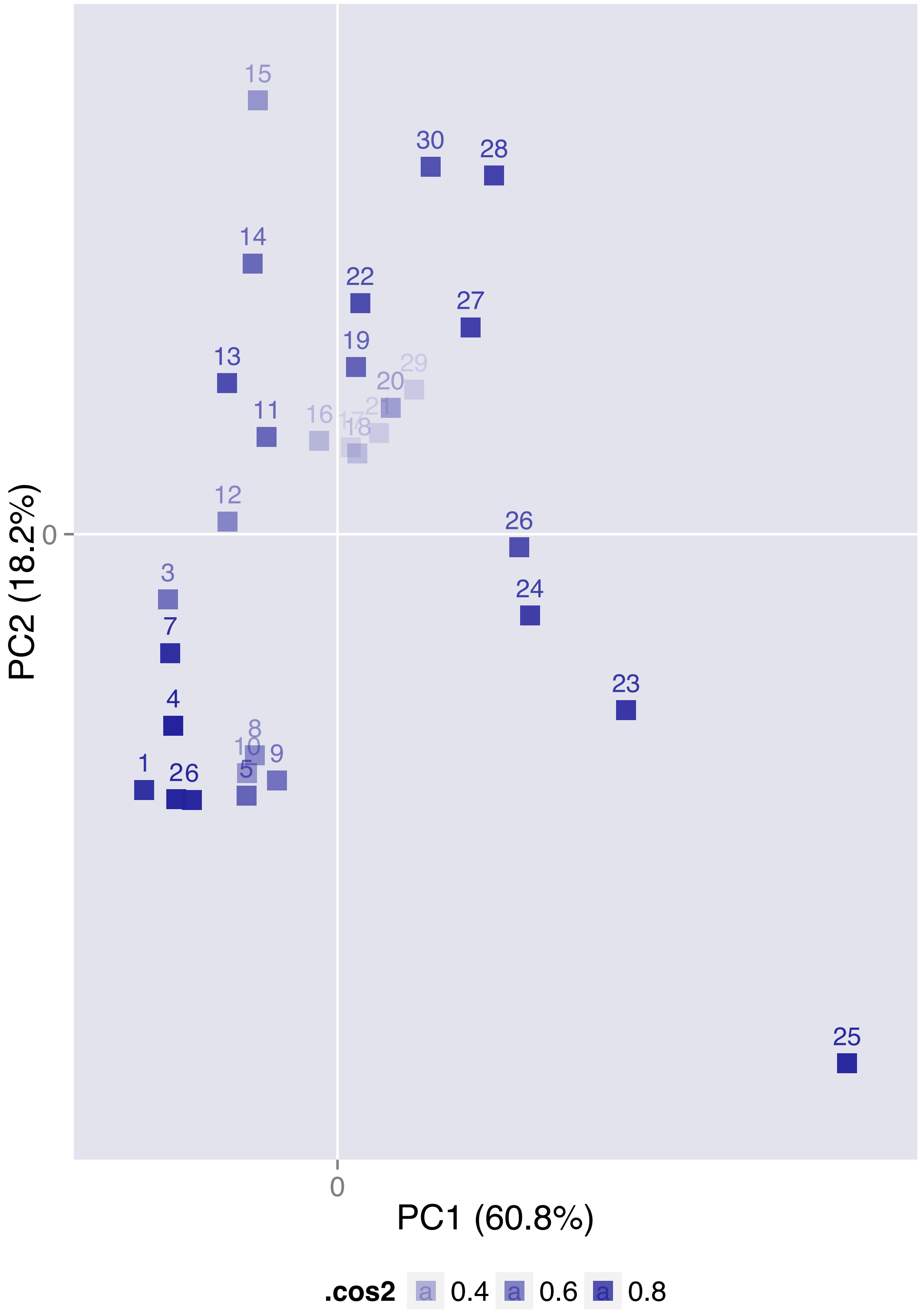
```
library("FactoMineR")

obj <- PCA(env, graph=F,
  quanti.sup=1, ind.sup=1)

autoplot(obj, type="var")

autoplot(obj, type="obs")
```



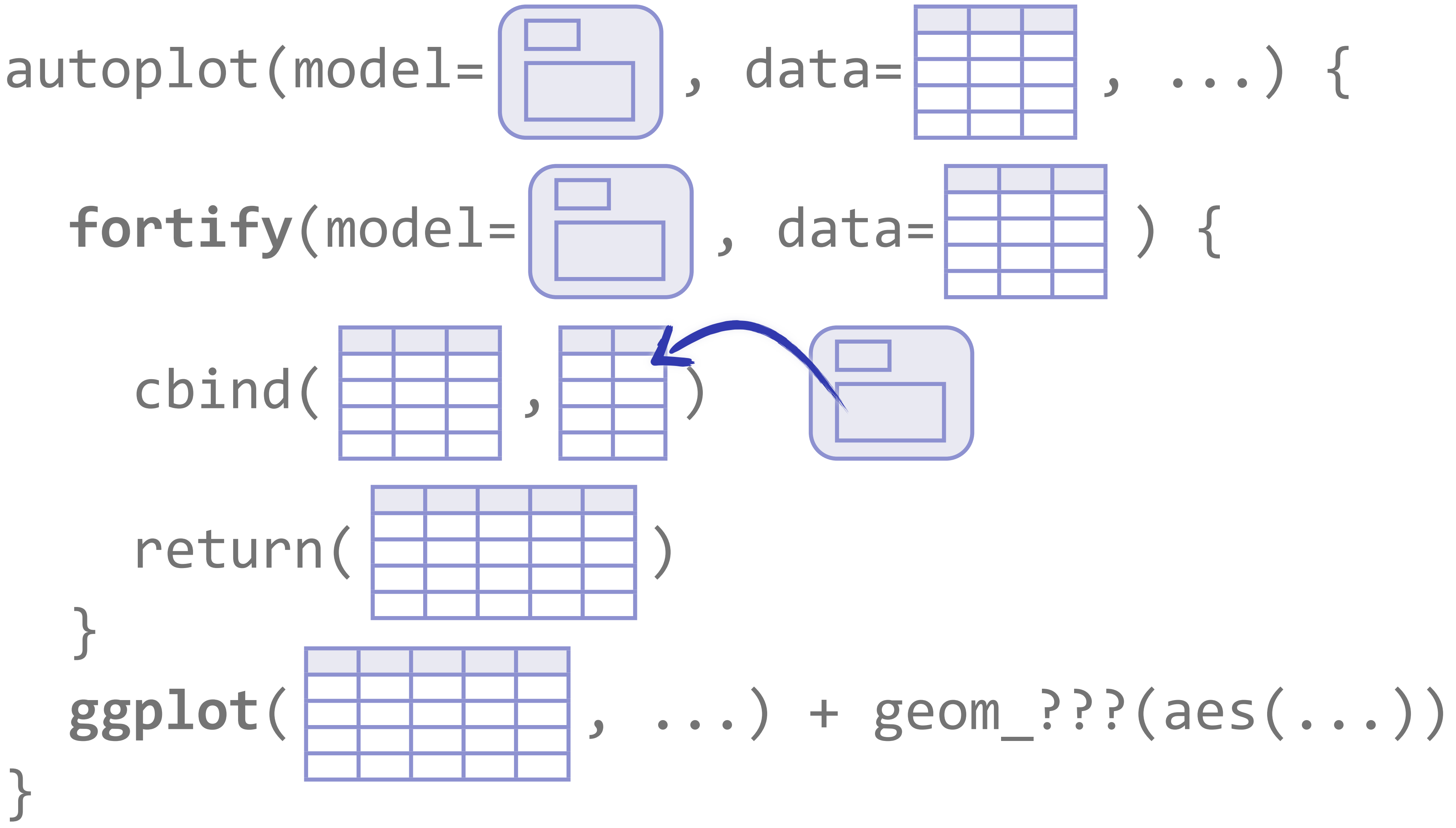


Yet customizable

```
obj <- PCA(env, graph=F)

p <- autoplot(obj, type="obs",

  # map additional variables
  mapping=aes(alpha=.cos2),

  # pass settings to ggplot calls
  shape=15, colour="#24239D"
)

# change the look after the fact
p + theme(
  panel.background=
  element_rect(fill="#E3E3EE"),
  legend.position="bottom"
)
# yay, beamer-like ggplot!
```



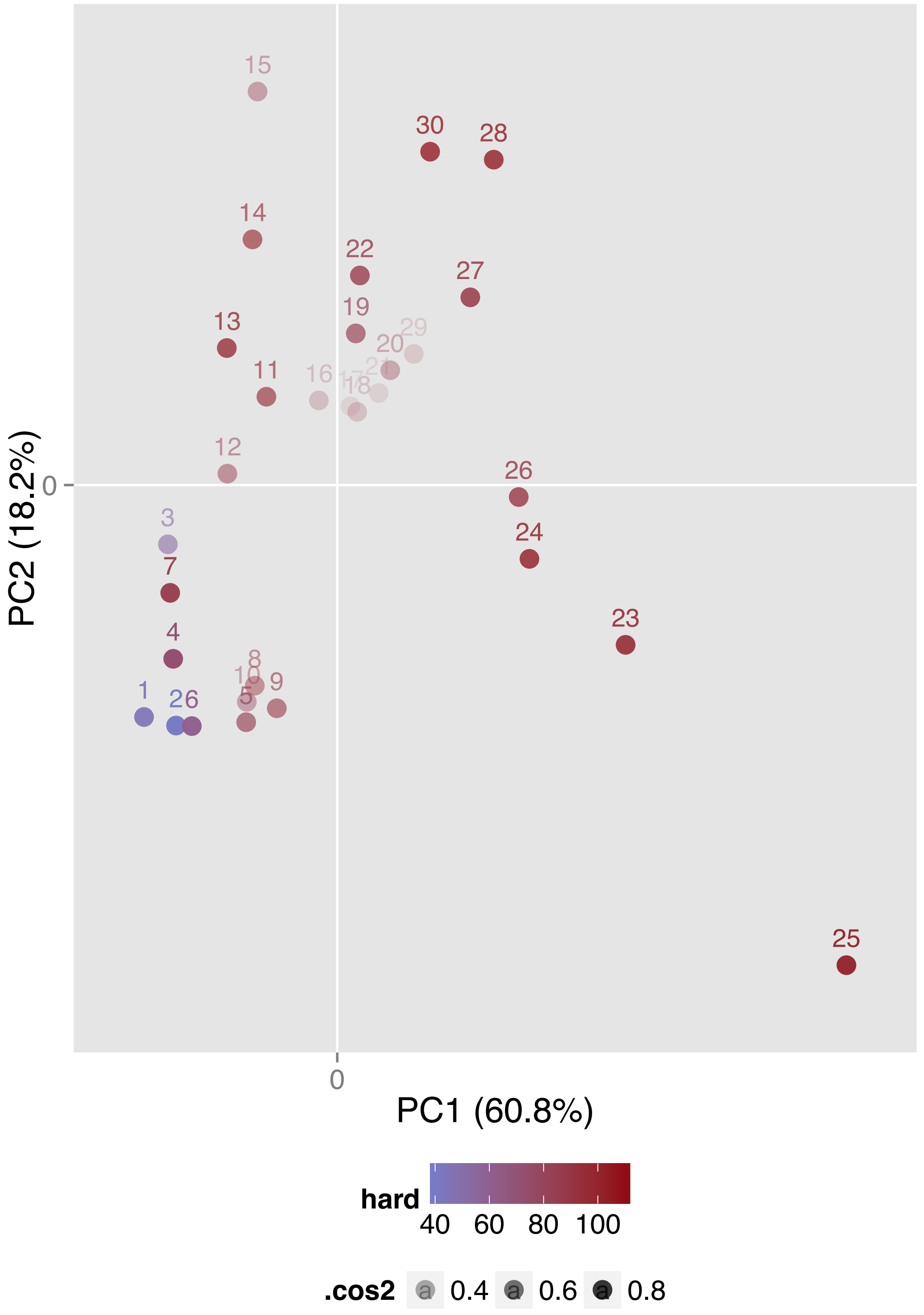
Two step process



```
autoplot(model=      , data=      , ...) {
   fortify(model=      , data=      ) {
      cbind(      ,      )
      return(      )
   }
   ggplot(      , ...) + geom_???(aes(...))
}
```

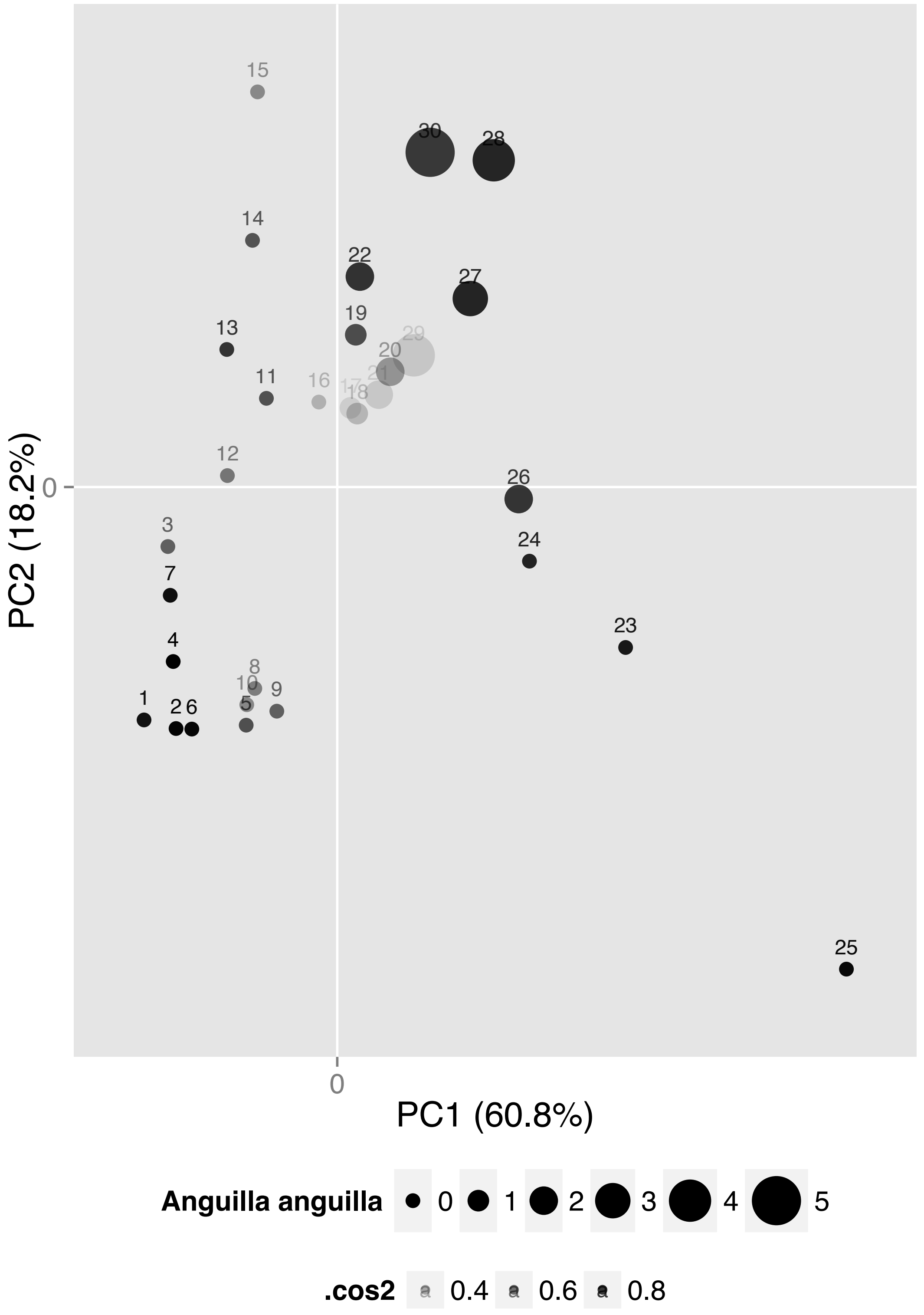
Map original variables, outside the analysis

```
head(env, 1)
  distance altitude pH phosphate
1        3      934 79         1
  nitrate ammonia  O2
1      20       0 122

head(allEnv, 1)
  distance altitude pH phosphate
1        3      934 79         1
  nitrate ammonia  O2 slope flow
1      20       0 122 6.176   84
  hard O2_demand
1   45        27

autoplot(obj, data=allEnv,
  type="obs",
  mapping=aes(alpha=.cos2,
    colour=hard)
)
```



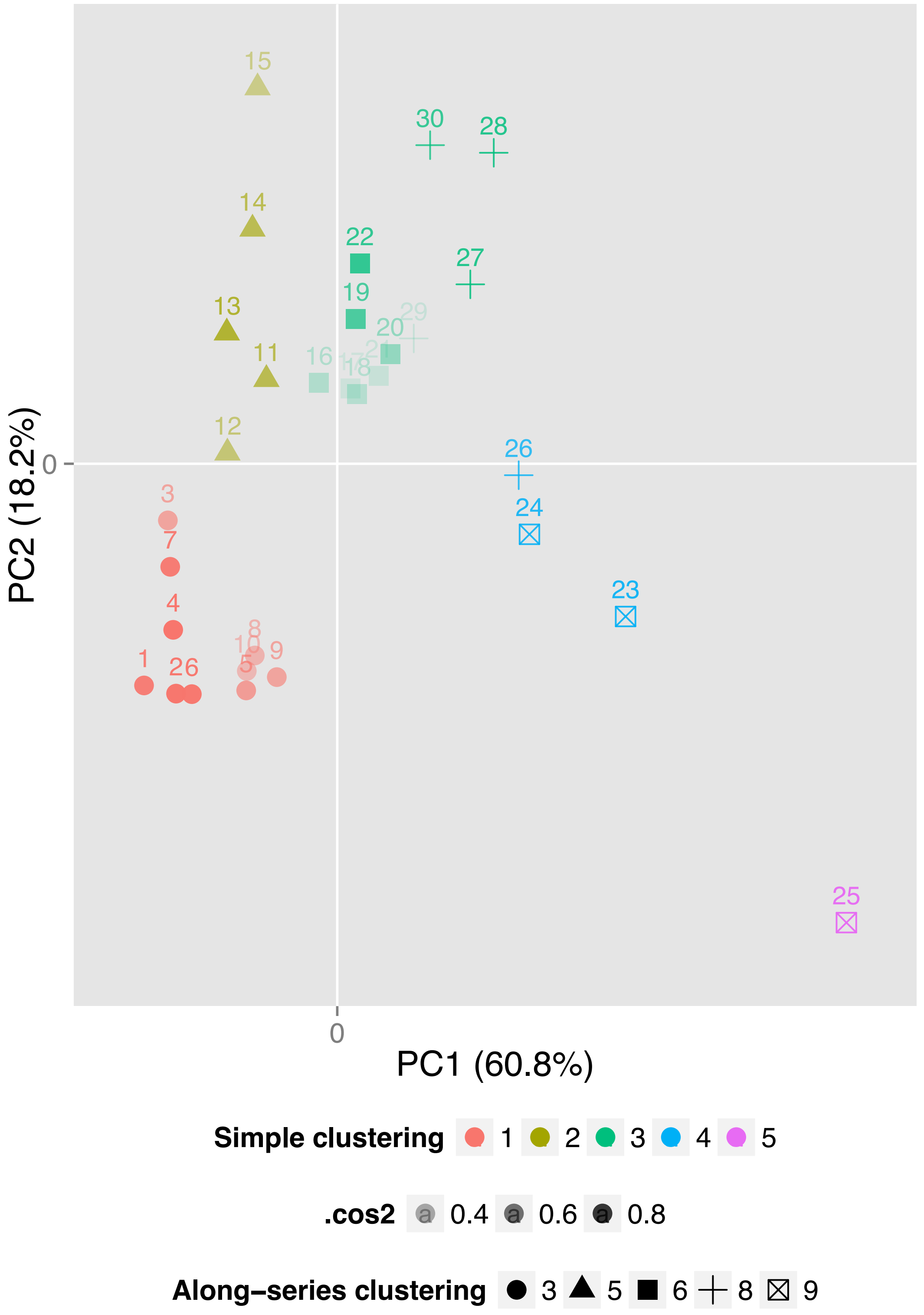
Map variables from other datasets

```
dim(env)
[1] 30  7

dim(fish)
[1] 30 27


autoplot(obj, data=fish,
  type="obs",
  mapping=aes(alpha=.cos2,
    size=`Anguilla anguilla`
  )
)
```



Easily visualize several results

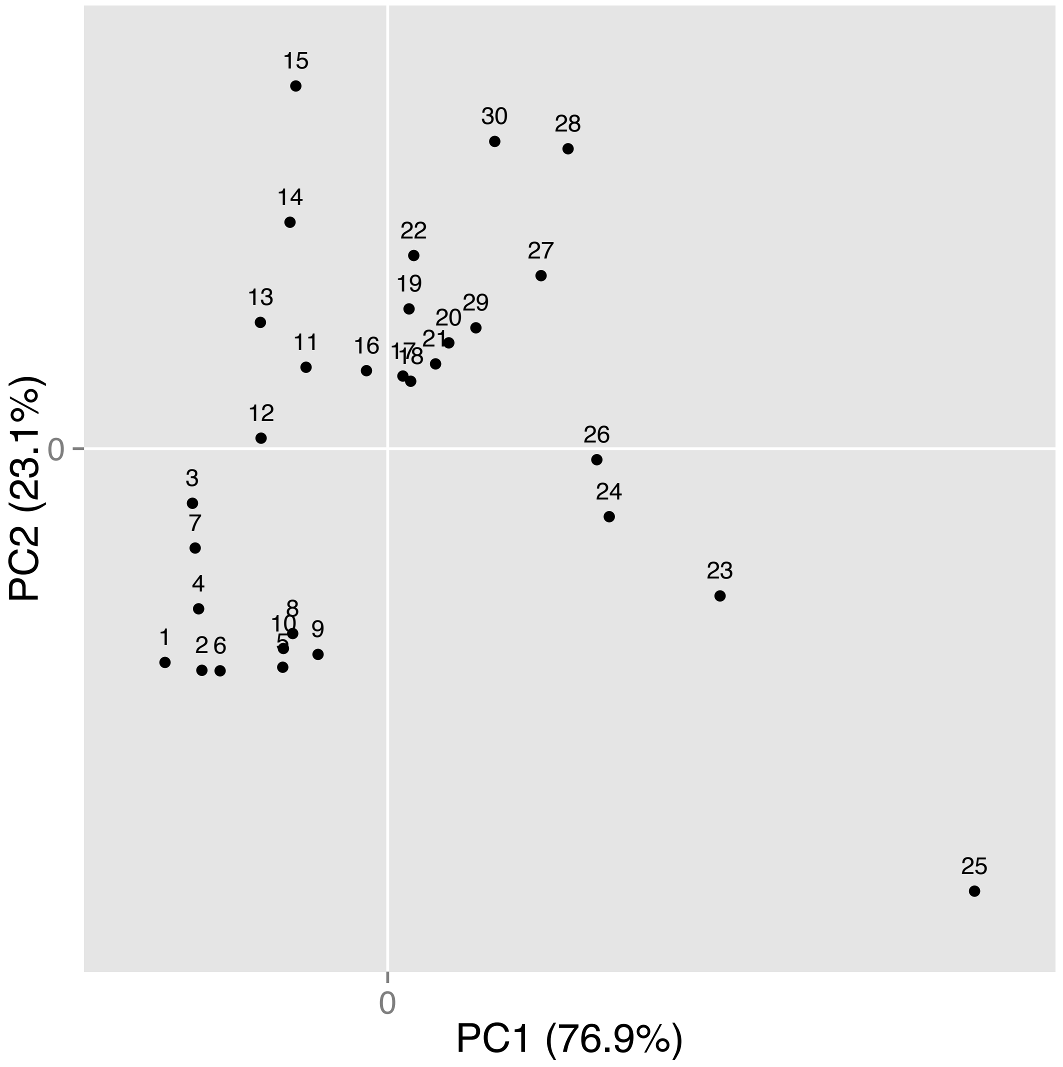
```
scores <- as.matrix(fortify(...

# un-constrainted clustering
clust <- hclust(
  dist(scores), method="ward")
env$clust <- cutree(clust, 5)
# clustering based on distance
library("mvpart")
dClust <- mvpart(
  scores ~ distance,
  data=env, size=5)
env$dclust <- dClust$where

autoplot(obj, data=env,
  type="obs",
  mapping=aes(alpha=.cos2,
    colour=clust,
    shape=dclust)
)
```



Homogenization at little cost

- **Several** fortify methods
- **One** plotting function

```
obj <- stats::prcomp(env,
  scale.=TRUE)
autoplot(obj, type="obs")

obj <- FactoMineR::PCA(env,
  graph=FALSE)
autoplot(obj, type="obs")

obj <- pcaMethods::pca(env,
  scale="uv")
autoplot(obj, type="obs")
```



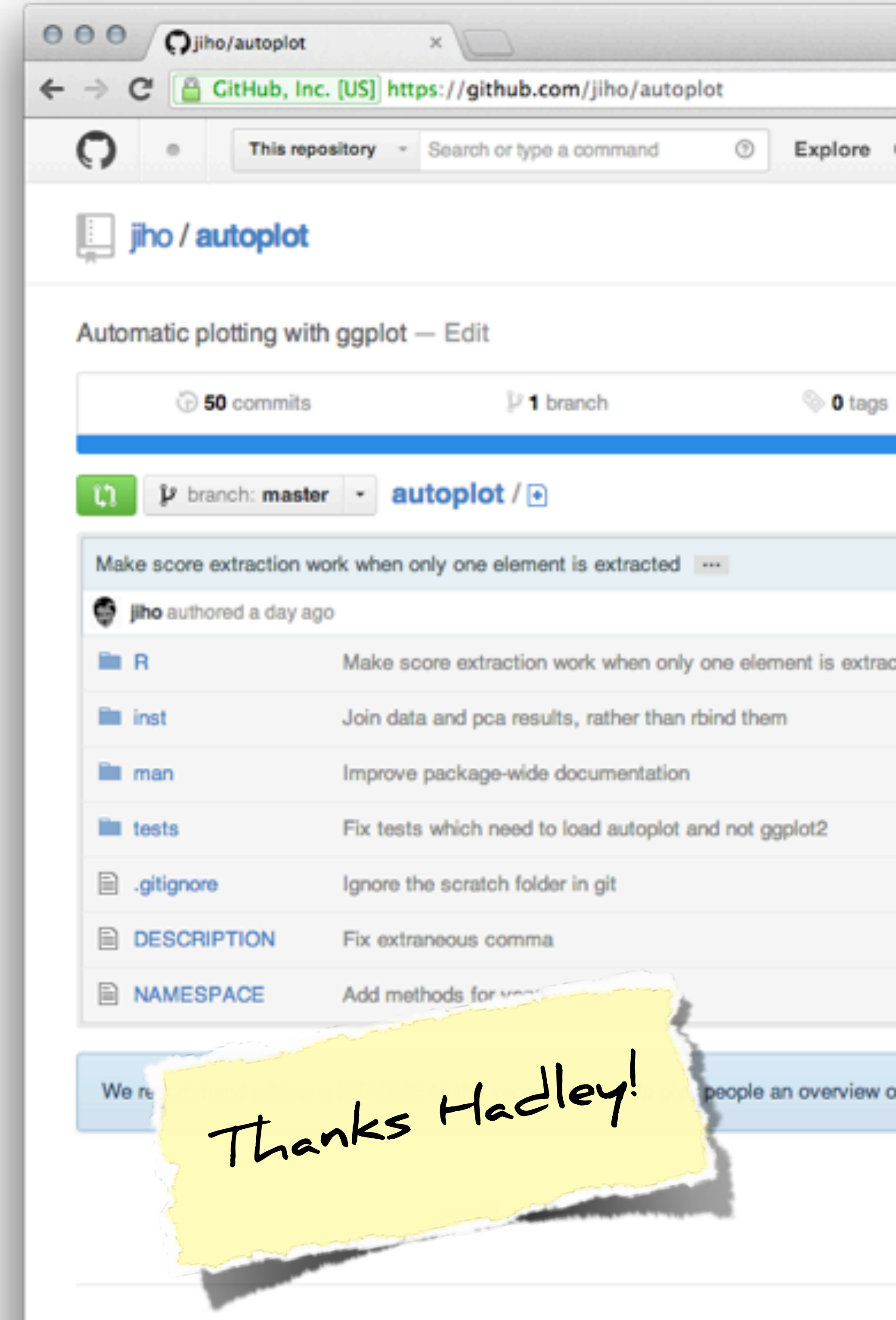
- **In progress:** `fortify.pca` (ade4) and `fortify.rda` (vegan)

Future directions

- Scaling + biplot in PCA
- More multivariate functions (CA, MCA are in progress)
- Code `autoplot.lm()` (`fortify.lm()` is already in ggplot2)
- Quantile regression (`rq`, `rqs`)
- Bayesian stats (`mcmc`)

- **https://github.com/jiho/autoplot**

```
library("devtools")
install_github("autoplot","jiho")
```



jiho/autoplot

GitHub, Inc. [US] https://github.com/jiho/autoplot

jiho / autoplot

Automatic plotting with ggplot — Edit

50 commits 1 branch 0 tags

branch: master autoplot /

Make score extraction work when only one element is extracted ...

jiho authored a day ago

R	Make score extraction work when only one element is extrac
inst	Join data and pca results, rather than rbind them
man	Improve package-wide documentation
tests	Fix tests which need to load autoplot and not ggplot2
.gitignore	Ignore the scratch folder in git
DESCRIPTION	Fix extraneous comma
NAMESPACE	Add methods for v

Thanks Hadley!