analysis_iv

Hunter York

10/31/2020

Examine how the import of skills has changed over time

This analysis uses the ONET 2016 skills file in conjunction with the ONET/DOT crosswalks. Data are the ACS 5-year samples from 2018, 2013, and 2009 (a slight overlap).

First, check how the skills rating themselves change over time, irrespective of census

```
# make a skills dataset that has a few interesting variables
skills_sum <- skills[Scale.ID == "LV",
                     .(average_value_skills = mean(Data.Value, na.rm = T)), by = .(year, OCCSOC)]
skills_sum[, year := as.character(year)]

# See which jobs are shared between all years
skills_sum[, keep := length(unique(year)) == 3, by = .(OCCSOC)]

skills_sum[, keep := T]

# compute average score by year
skills_by_year <- skills[Scale.ID == "LV",
                     .(average_value_skills = mean(Data.Value, na.rm = T)), by = .(year)]
skills_by_year
```

```
##    year average_value_skills
## 1: 2009            0.4980616
## 2: 2013            0.4829824
## 3: 2018            0.4812640
```

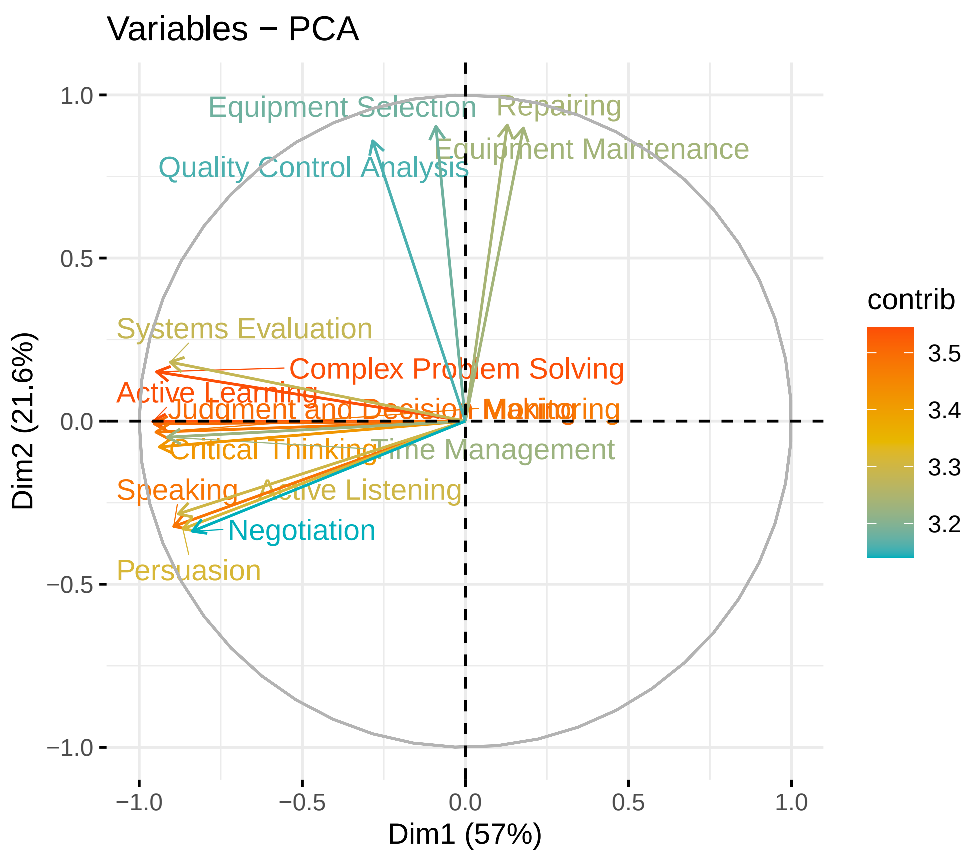
```
# average across years
# REMOVE THIS IF YOU WANT YEAR SPECIFIC SKILLS RATINGS
skills[,Data.Value := mean(Data.Value), by = .(Element.Name, Scale.ID, OCCSOC)]
```

We notice that there's a decrease in the average skill level between 2009 and 2013, even after using percentil standardizations.

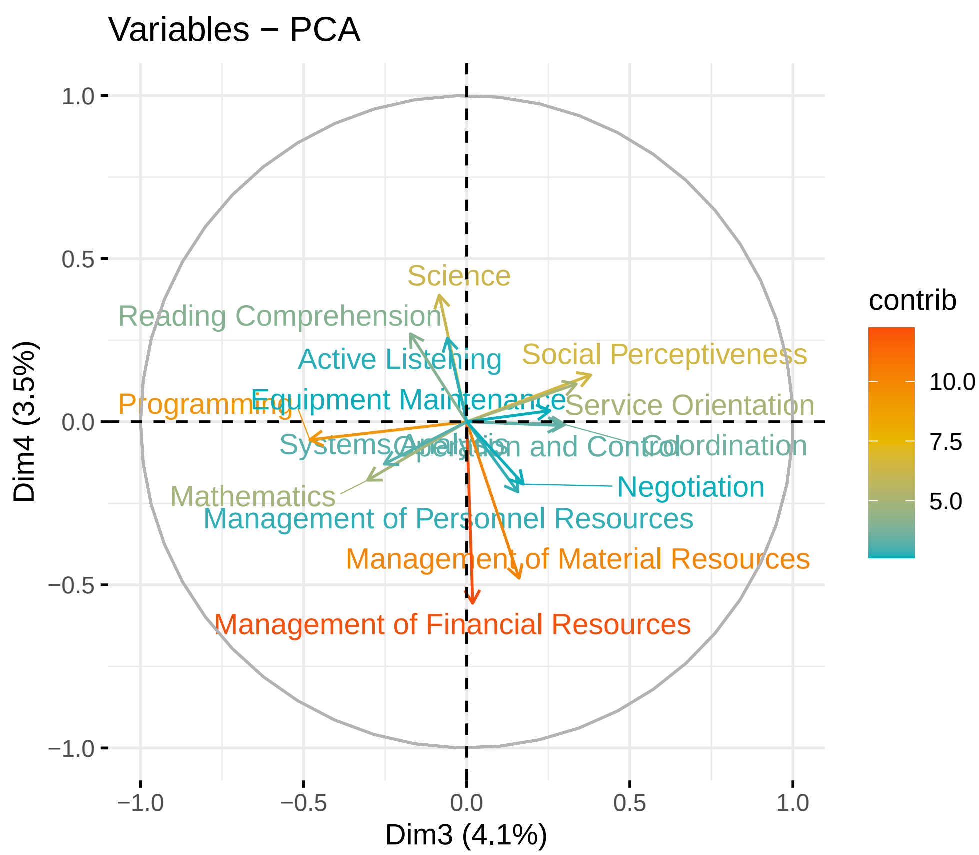
Do a quick PCA to see how skills vary with respect to each other

```
# cast wide and do factor analysis
skills <- skills[OCCSOC %in% unique(acs$OCCSOC),
                 .(Data.Value = mean(Data.Value)), by = .(Element.Name, Scale.ID, year, OCCSOC)]
```

```r
skills_wide <- dcast(skills[Scale.ID == "LV"], OCCSOC  + year ~Element.Name, value.var = "Data.Value")
res.pca <- prcomp(skills_wide[,3:35], scale = TRUE, center = T)

fviz_pca_var(res.pca,
             col.var = "contrib", # Color by contributions to the PC
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE ,
             select.var = list(contrib = 15)     # Avoid text overlapping
)
```

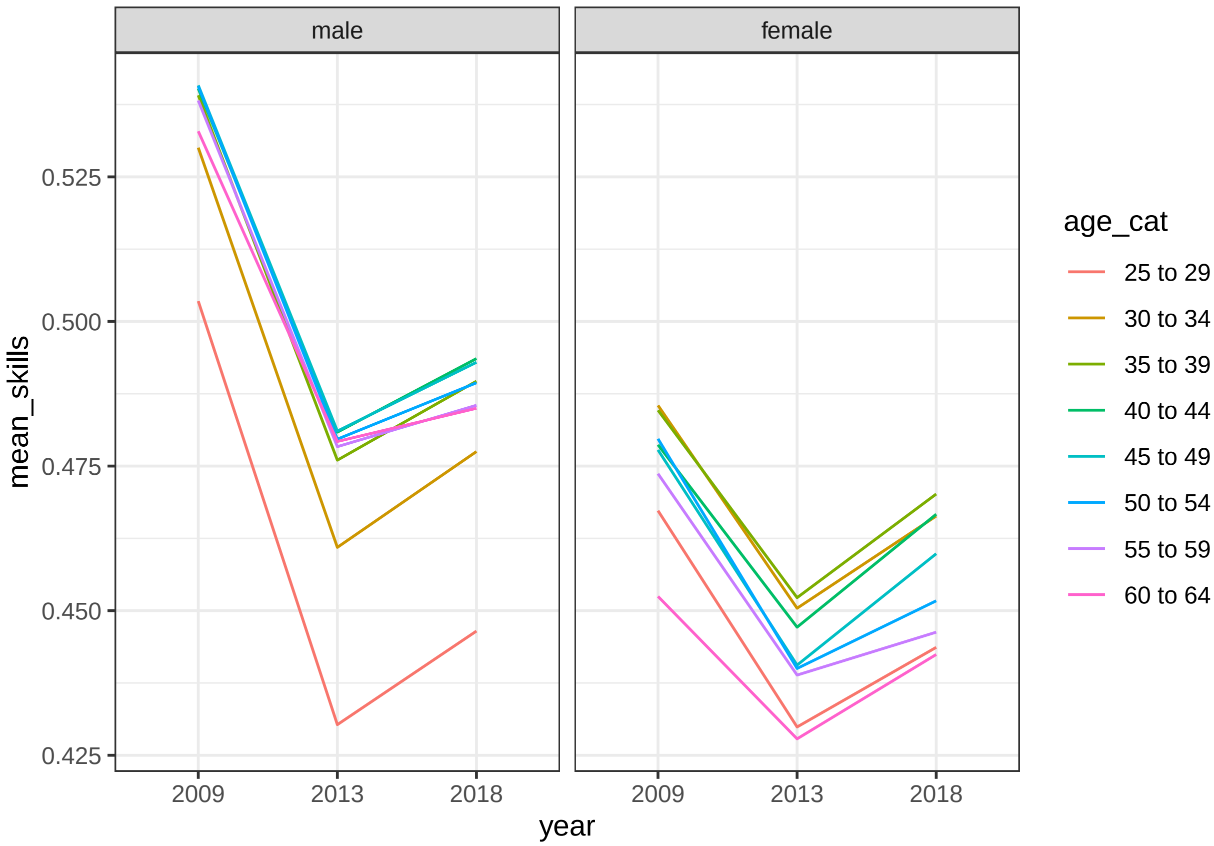



```r
fviz_pca_var(res.pca,
             axes = c(3,4),
             col.var = "contrib", # Color by contributions to the PC
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE,# Avoid text overlapping,
             select.var = list(contrib = 15)
)
```

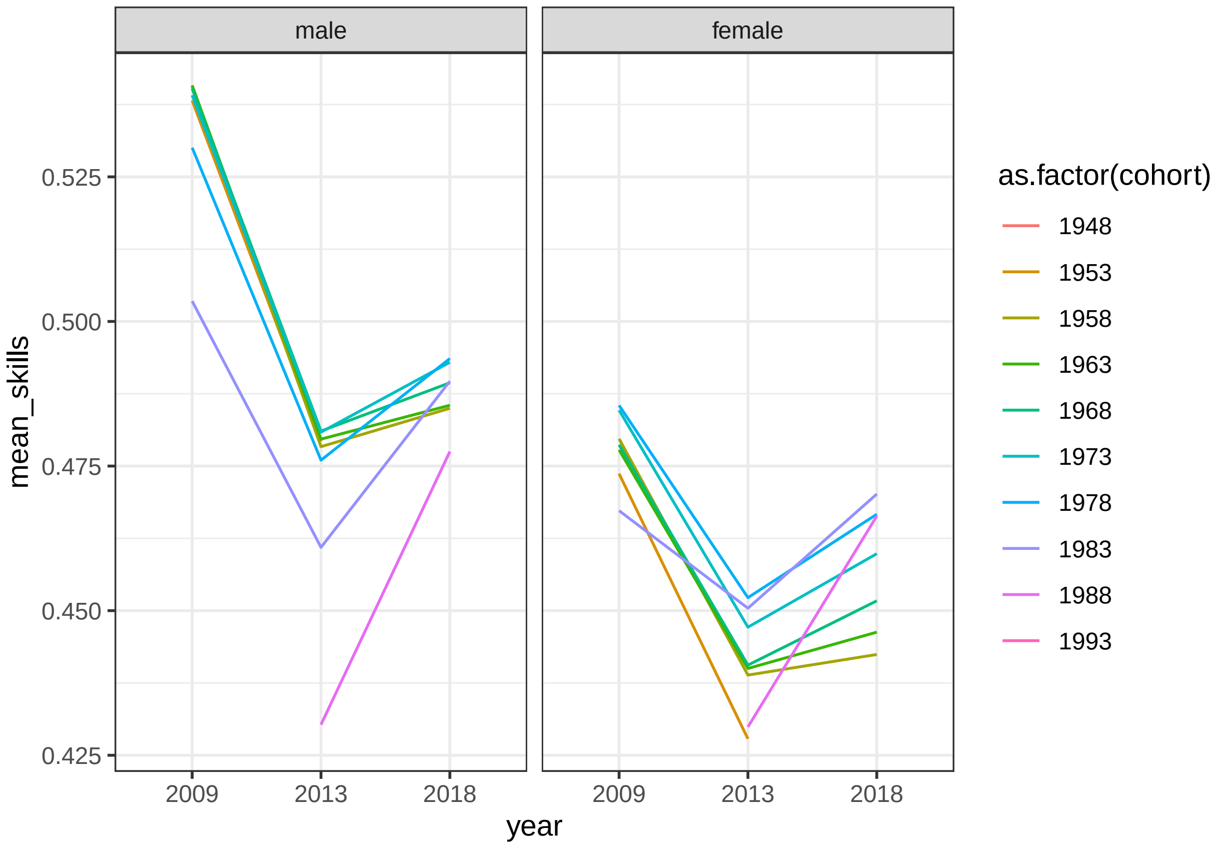


See how mean skills change across survey year using ACS data

```
acs <- merge(acs, skills_sum, by = c("year", "OCCSOC"), all.x = T)

skills_overview <- acs[,.(mean_skills = weighted.mean(average_value_skills, w = perwt, na.rm = T)), by =

skills_overview[year > 2009, cohort := as.numeric(year) - as.numeric(substr(age_cat,1,2))]
skills_overview[year == 2009, cohort := as.numeric(year)-1 - as.numeric(substr(age_cat,1,2))]

ggplot(skills_overview) +
    geom_line(aes(x = year, y = mean_skills, color = age_cat, group = age_cat)) +
    facet_wrap(~sex)
```




```r
ggplot(skills_overview) +
    geom_line(aes(x = year, y = mean_skills, color = as.factor(cohort), group = as.factor(cohort))) +
    facet_wrap(~sex)
```



Recalculate skills by year to capture more interesting variables, like LV1 and LV2, Programming, etc

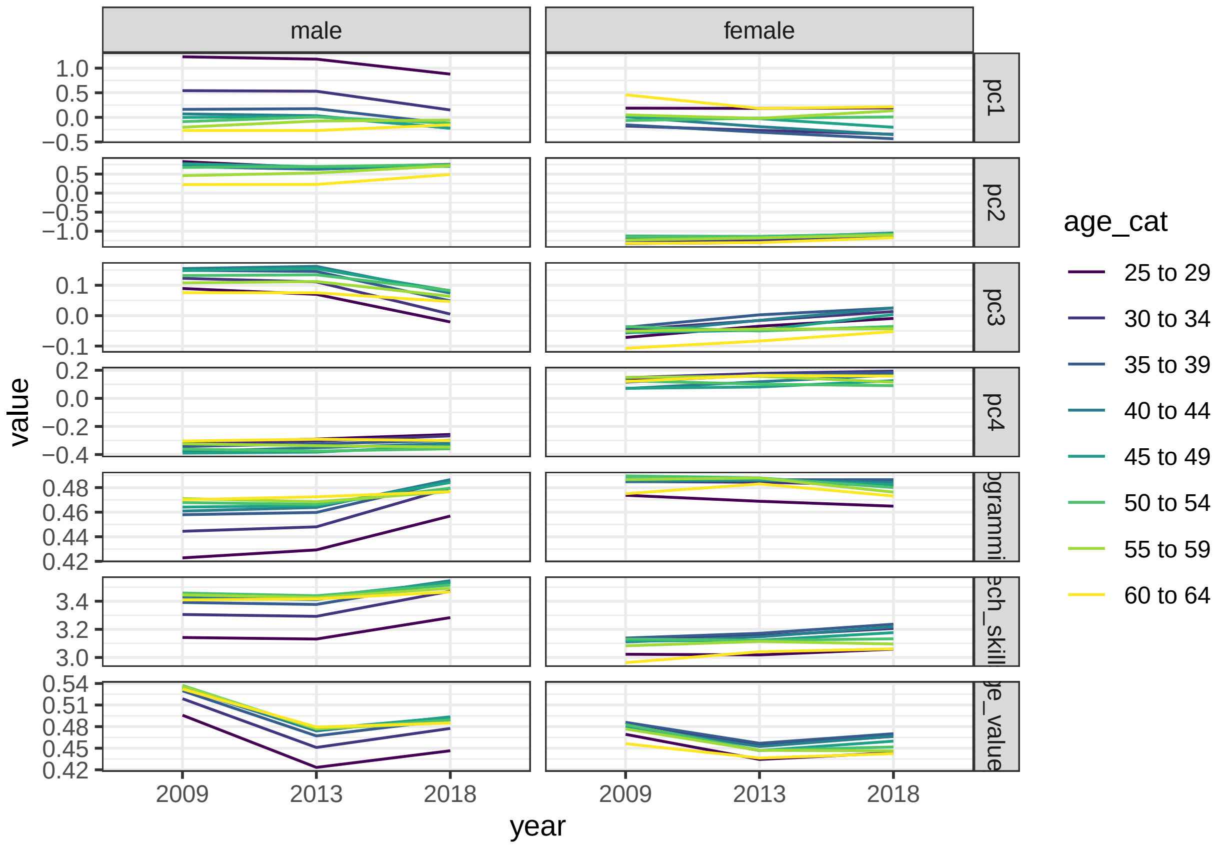
```r
# skills wide
skills_sum <- skills_wide
skills_sum[, pc1 := predict(res.pca, newdata = .SD)[,1], .SDcols = names(skills_sum)]
skills_sum[, pc2 := predict(res.pca, newdata = .SD)[,2], .SDcols = names(skills_sum)]
skills_sum[, pc3 := predict(res.pca, newdata = .SD)[,3], .SDcols = names(skills_sum)]
skills_sum[, pc4 := predict(res.pca, newdata = .SD)[,4], .SDcols = names(skills_sum)]

skills_sum[, programming := Programming]
skills_sum[, tech_skills := Programming + `Complex Problem Solving` +
               `Mathematics` + Programming + Science + `Systems Analysis` +
               Troubleshooting]
skills_sum[, average_value_skills := rowMeans(.SD), .SDcols = rownames(res.pca$rotation)]

skills_sum[, year := as.character(year)]

# See which jobs are shared between all years
skills_sum[, keep := length(unique(year)) == 3, by = .(OCCSOC)]

# compute average score by year
skills_by_year <- skills_sum[keep == T,.(average = mean(average_value_skills),
                                         pc1 = mean(pc1),
                                         pc2 = mean(pc2),
                                         pc3 = mean(pc3),
```

```
                                        pc4 = mean(pc4),
                                        programming = mean(programming),
                                        tech_skills = mean(tech_skills)), by = year]
skills_by_year
```

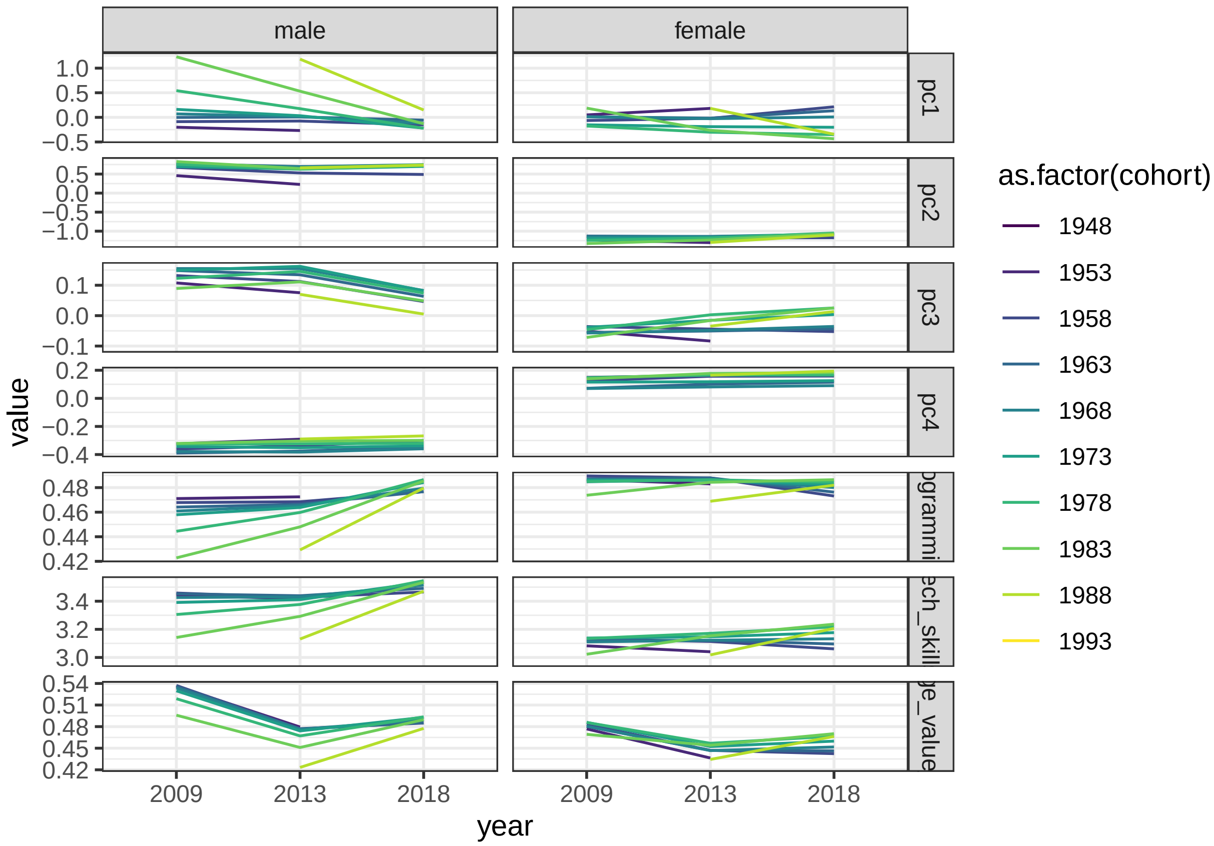
```
##    year   average          pc1           pc2          pc3           pc4
## 1: 2009 0.4793571 4.200844e-17 -2.400482e-17 9.414391e-17 -2.385479e-16
## 2: 2013 0.4793571 4.200844e-17 -2.400482e-17 9.414391e-17 -2.385479e-16
## 3: 2018 0.4793571 4.200844e-17 -2.400482e-17 9.414391e-17 -2.385479e-16
##    programming tech_skills
## 1:   0.4703533    3.347141
## 2:   0.4703533    3.347141
## 3:   0.4703533    3.347141
```

See how mean skills change across survey year

```
# average across years
# REMOVE THIS IF YOU WANT YEAR SPECIFIC SKILLS RATINGS
# skills_sum <- skills_sum[year == 2018,.(pc1 = mean(pc1),
#                             pc2 = mean(pc2),
#                             programming = mean(programming)), by = "OCCSOC"]

acs <- merge(acs, skills_sum[,.(OCCSOC, year, pc1, pc2,pc3, pc4, programming, tech_skills)], by = c( "OC

# only retain jobs that are in common between all three years

skills_overview2 <- acs[OCCSOC %in% acs[!is.na(pc1) & year == 2018, unique(OCCSOC)],.(pc1 = weighted.mea
                                                                                      pc2 = weighted.mea
                                                                                      pc3 = weighted.mea
                                                                                      pc4 = weighted.mea
                                                                                      programming = weig
                                                                                      tech_skills = weig
                                                                                      average_value_skil

skills_overview2_melt <- melt(skills_overview2, id.vars = c("year", "age_cat", "sex"))

skills_overview2_melt[year > 2009, cohort := as.numeric(year) - as.numeric(substr(age_cat,1,2))]
skills_overview2_melt[year == 2009, cohort := as.numeric(year)-1 - as.numeric(substr(age_cat,1,2))]


ggplot(skills_overview2_melt) +
    geom_line(aes(x = year, y = value, color = age_cat, group = age_cat)) +
    facet_grid(variable~sex, scales = "free") +
    scale_color_viridis_d()
```

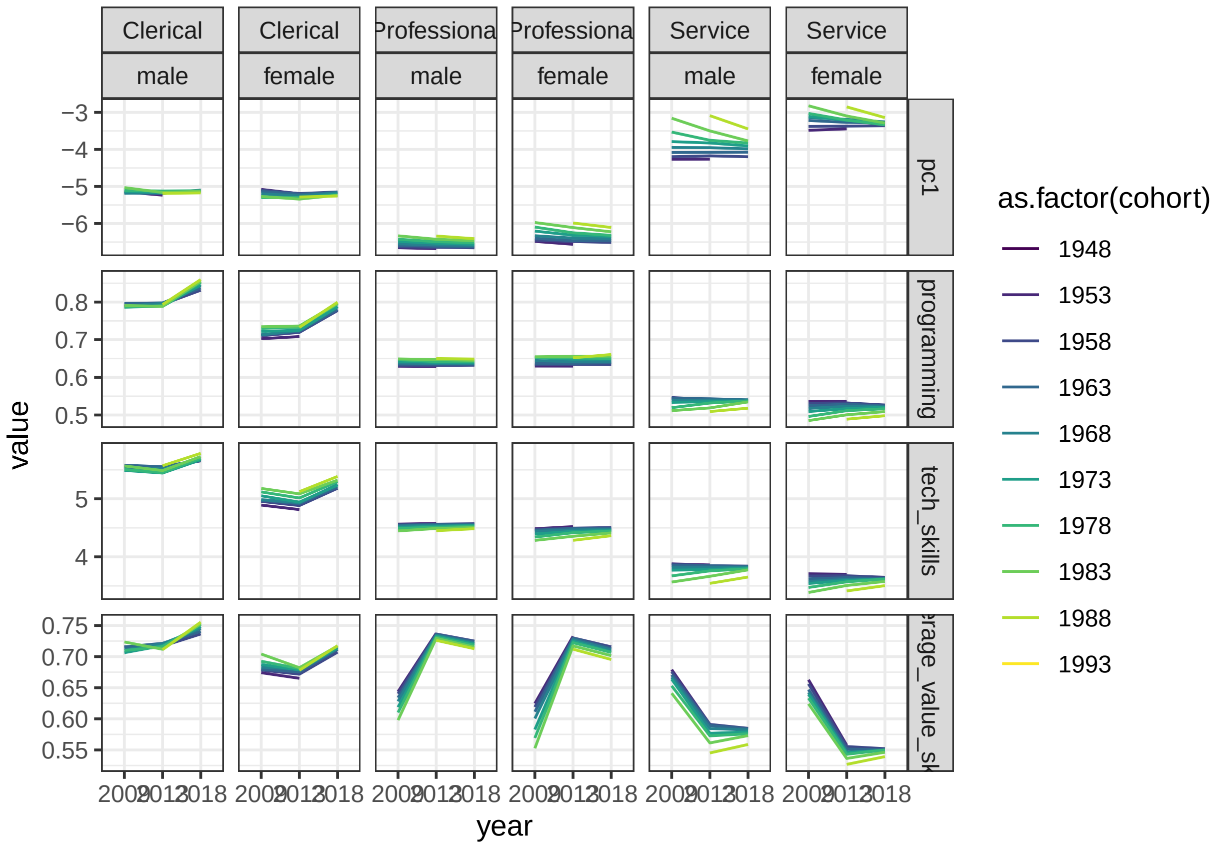



```
ggplot(skills_overview2_melt) +
    geom_line(aes(x = year, y = value, color = as.factor(cohort), group = as.factor(cohort))) +
    facet_grid(variable~sex, scales = "free") +
    scale_color_viridis_d()
```

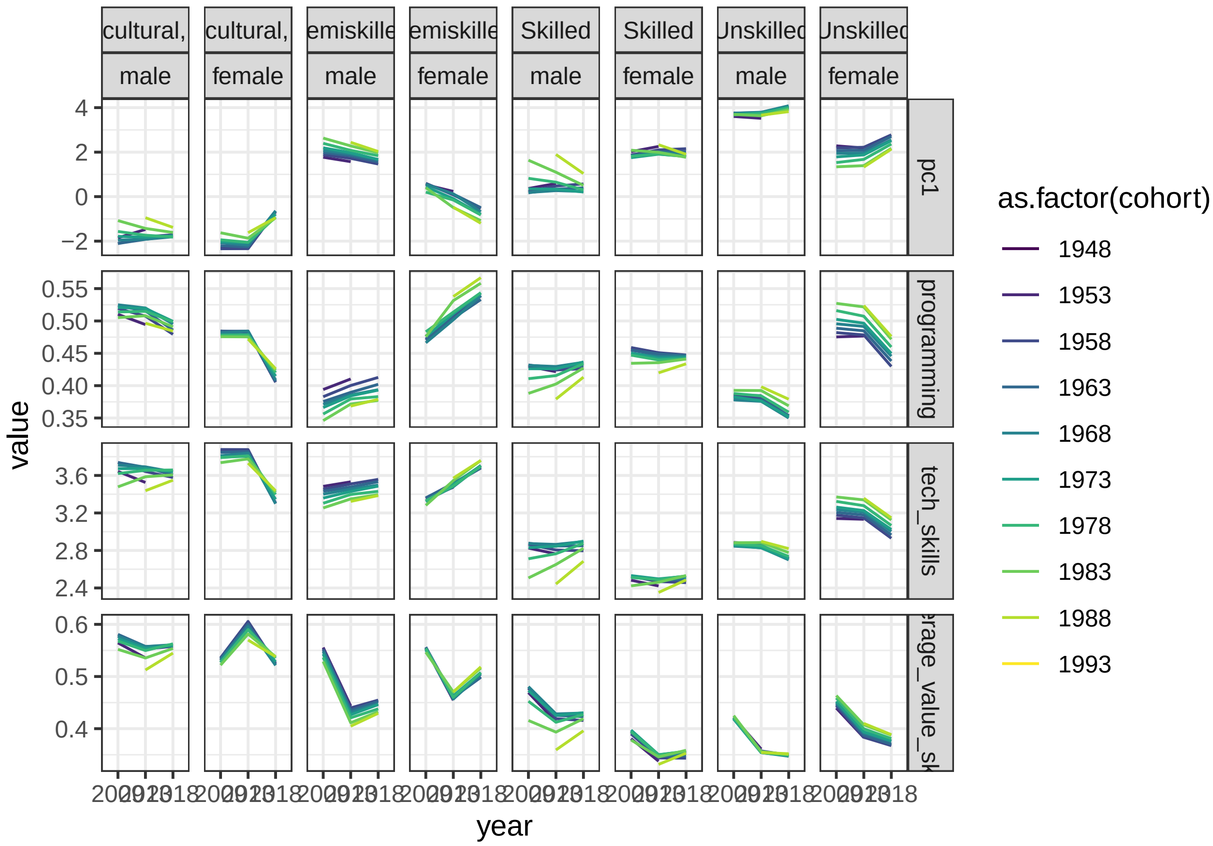


See how changes in skill level are spread over occupational grouping

```
skills_overview3 <- acs[OCCSOC %in% acs[!is.na(pc1) & year == 2018, unique(OCCSOC)],.(pc1 = weighted.mea
                                                                                        pc2 = weighted.mea
                                                                                        pc3 = weighted.mea
                                                                                        pc4 = weighted.mea
                                                                                        programming = weig
                                                                                        tech_skills = weig

skills_overview3_melt <- melt(skills_overview3, id.vars = c("year", "age_cat", "sex" ,"occ_categ"))
skills_overview3_melt[year > 2009, cohort := as.numeric(year) - as.numeric(substr(age_cat,1,2))]
skills_overview3_melt[year == 2009, cohort := as.numeric(year)-1 - as.numeric(substr(age_cat,1,2))]


ggplot(skills_overview3_melt[variable %like% "pc1|progr|tech|averag" &
                                  occ_categ %in% unique(skills_overview3_melt$occ_categ)[1:3]]) +
    geom_line(aes(x = year, y = value, color = as.factor(cohort), group = as.factor(cohort))) +
    facet_grid( variable ~occ_categ+sex, scales = "free") +
    scale_color_viridis_d()
```




```
ggplot(skills_overview3_melt[variable %like% "pc1|progr|tech|averag" &
                                  occ_categ %in% unique(skills_overview3_melt$occ_categ)[4:7]]) +
    geom_line(aes(x = year, y = value, color = as.factor(cohort), group = as.factor(cohort))) +
    facet_grid( variable ~occ_categ+sex, scales = "free") +
    scale_color_viridis_d()
```



The fact that age groups vary consistently across ACS years indicates that year changes are artifacts of the data. b/w cohort changes are interesting, however.

Examine how class of worker, industry, and ed affect inequality