## Analyzing the Most Recent OSU Salary Report (2022)

```
source("salary_pdf_parser.r")

fname <- "_data_unclassifieds/unclassified_2022-04-18.pdf"
employees <- parse_salaries_single_year(fname)
```

### Q1: How do the distributions of salary compare between ranks of teachers?

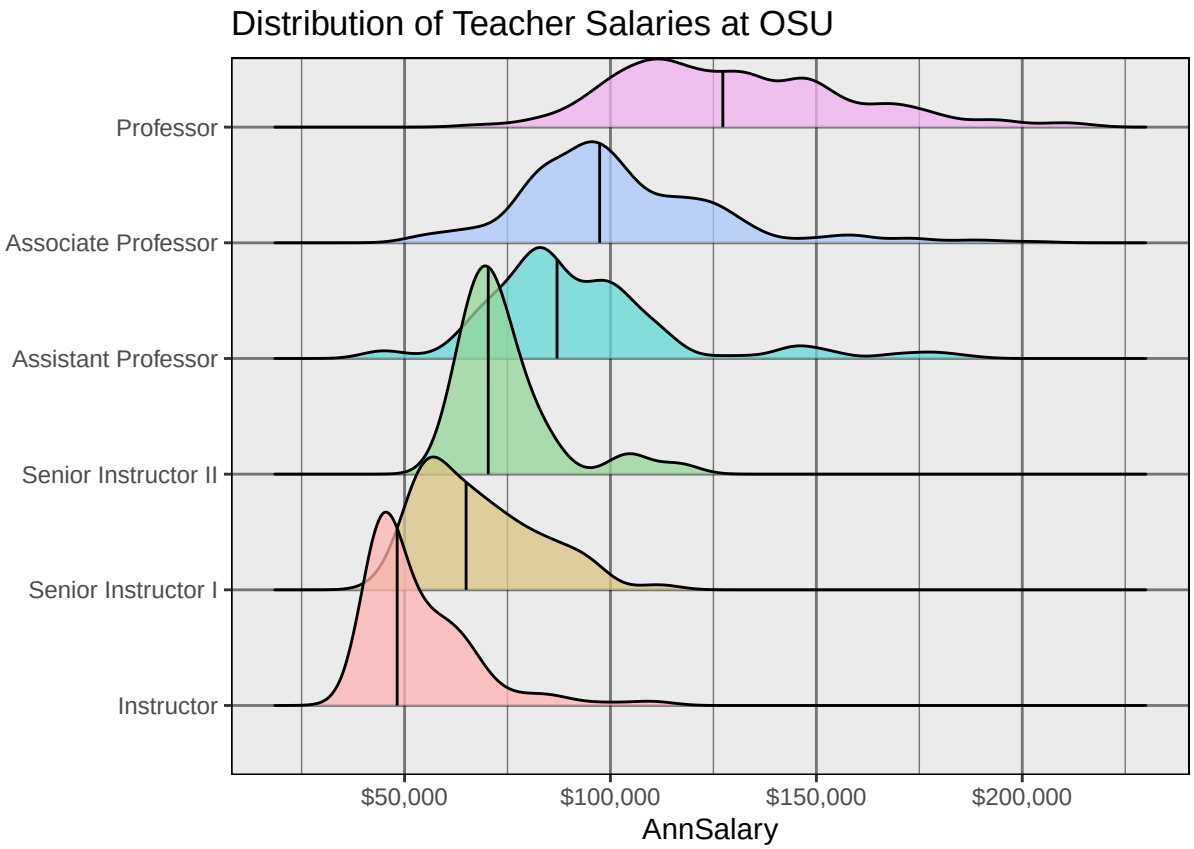
```
library(ggridges)

my_jobs <- c("Instructor", "Senior Instructor I", "Senior Instructor II",
             "Assistant Professor", "Associate Professor", "Professor")
my_job_regex <- paste("(^", my_jobs, "$)", sep="", collapse="|")

employees.teachers <- (
    employees
    %>% drop_na(AnnSalary)
    %>% mutate(AnnSalary = ifelse(ApptType==12, AnnSalary*(0.75), AnnSalary))
    %>% filter(str_detect(JobTitle, my_job_regex))
    %>% mutate(JobTitle = factor(JobTitle, levels=my_jobs, ordered=TRUE))
)

### Note: these are not ALL instructors/professors. Some have been left out
###       because their title has an extra part (e.g. "Professor (Clinical)")
### Note: these are SCALED salaries, i.e. those with 12-month appointments
###       have theirs scaled by (3/4) to be on the same scale as 9-month appts

ggplot(employees.teachers) +
    geom_density_ridges(aes(x=AnnSalary, y=JobTitle, fill=JobTitle),
                        alpha=0.75,
                        quantile_lines=TRUE,
                        quantiles=c(0.5)) +
    scale_x_continuous(breaks=1000*seq(50,225,50),
                       labels=scales::label_dollar()) +
    scale_fill_manual(values=hcl.colors(6, palette="Set3")) +
    labs(title = "Distribution of Teacher Salaries at OSU", y = "") +
    theme(panel.grid = element_line(color=scales::alpha("black", 0.5)),
          panel.border=element_rect(color="black", fill=NA),
          legend.position="none")
```

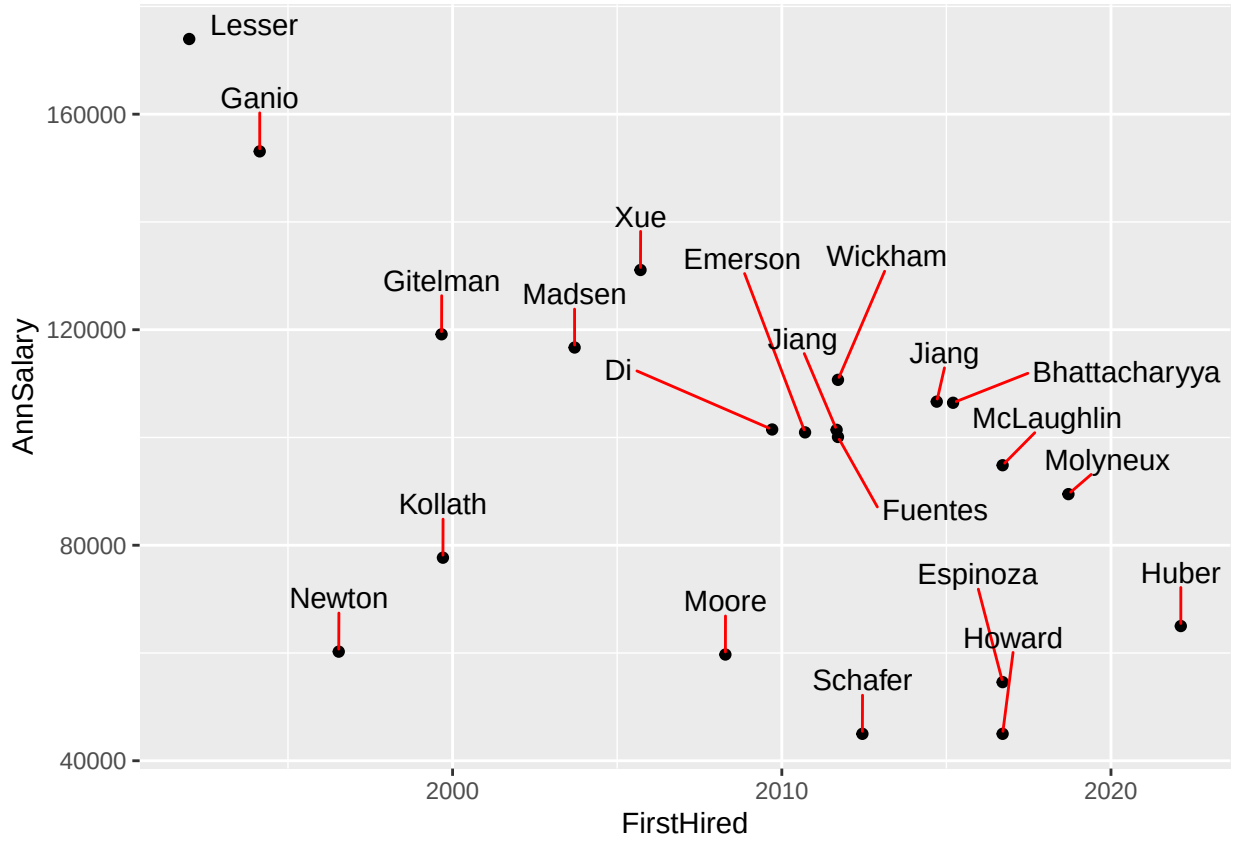

Distribution of Teacher Salaries at OSU
Professor
Associate Professor
Assistant Professor
Senior Instructor II
Senior Instructor I
Instructor
$50,000
$100,000
$150,000
$200,000
AnnSalary

**Q2: How do salaries compare within the statistics department?**

```
library(ggrepel)

employees.stats <- (
    employees
    %>% filter(str_detect(JobOrgn, "Statistics"))
    %>% filter(str_detect(JobTitle, "(Emeritus)|(Courtesy)", negate=TRUE))
)

ggplot(employees.stats, aes(x=FirstHired, y=AnnSalary)) +
    geom_point() +
    geom_text_repel(aes(label=LastName), seed=1, force=50,
                    nudge_x = 1, nudge_y = 10000,
                    segment.color="red")
```



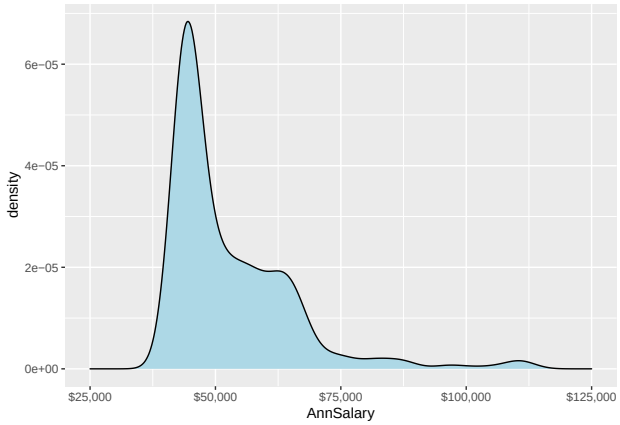
**Q3: What is the distribution of salary for JUST instructors on 9mo appts?**

```
instructors.9mo <- (
    employees
    %>% drop_na(AnnSalary)
    %>% filter(JobTitle == "Instructor", ApptType == 9)
)

instructors.9mo %>% use_series(AnnSalary) %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   42993   44298   47651   52842   58815  112005
```

```
ggplot(instructors.9mo) +
    geom_density(aes(x=AnnSalary), fill="lightblue") +
    scale_x_continuous(breaks=seq(0, 125000, 25000),
                       labels=scales::label_dollar(),
                       limits=c(25000, 125000))
```




```
#########################################################
### Who are the instructors making > 100K salary? ###
#########################################################

instructors.9mo %>%
    filter(AnnSalary >= 100000) %>%
    select(LastName, FirstHired, JobOrgn, AnnSalary)
```

```
##          LastName FirstHired                      JobOrgn AnnSalary
## 1           Clark 2010-09-16       CLA - Music Department    110925
## 2           Garza 2021-09-16       CLA - Music Department    104526
## 3          Hebert 2017-09-16       CLA - Music Department    110853
## 4 Kosanovic-Brown 2010-09-16       CLA - Music Department    110925
## 5        Mansouri 2009-01-16 VBS - Vet Biomedical Science    112005
## 6          Reneau 2019-09-16       CLA - Music Department    107658
## 7         Talcott 2020-03-16 VBS - Vet Biomedical Science    110016
## 8           White 2019-10-01 VBS - Vet Biomedical Science    100008
```

## Analyzing OSU Salary Reports Since 2014

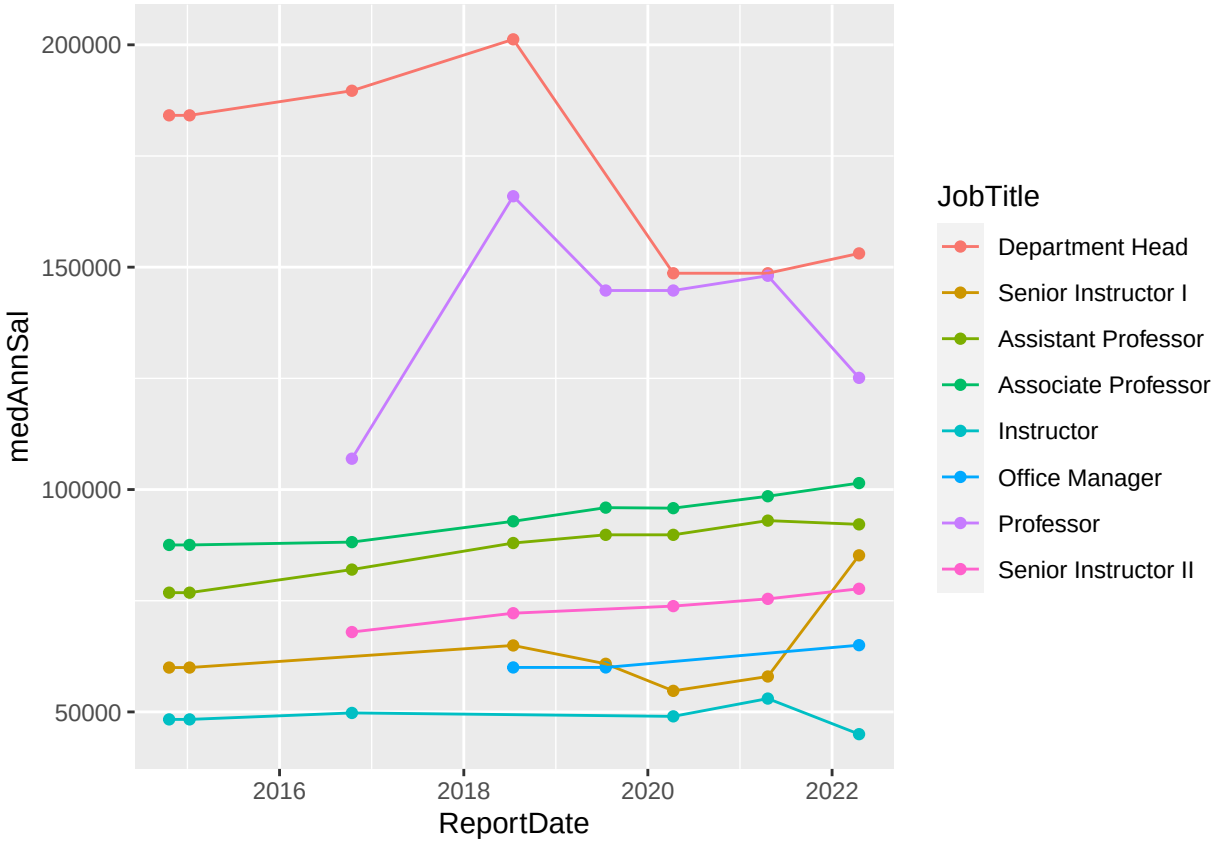
```
rm(list = ls()) # clear the environment

# source("salary_pdf_parser.r")
#
# fnames <- list.files("_data_unclassifieds", ".*\\.pdf") %>%
#             { paste("_data_unclassifieds/", ., sep="") }
# employees.all <- parse_salaries_all_years(fnames)
#
# saveRDS(employees.all, file="_Unclassifieds_All_Years.rds")

employees.all <- readRDS("_Unclassifieds_All_Years.rds")
```

### Q1: How has Median Salary Changed WITHIN the Stats Dept?

```
employees.all %>%
    drop_na(AnnSalary) %>%
    filter(str_detect(JobOrgn, "Statistics")) %>%
    filter(!str_detect(JobTitle, "(Emeritus)|(Courtesy)")) %>%
    filter(!str_detect(JobTitle, "(Research)")) %>%
    mutate(
        JobTitle = recode_factor(
                        JobTitle,
                        `Department Chair` = "Department Head",
                        `Department Head -Statistics` = "Department Head",
                        `Senior Instructor 1` = "Senior Instructor I"
                    )
    ) %>%
    group_by(ReportDate, JobTitle) %>%
    summarize(medAnnSal = median(AnnSalary)) %>%
    ungroup() %>%
    ggplot(aes(x=ReportDate, y=medAnnSal, color=JobTitle)) +
        geom_point() +
        geom_line()
```


200000
150000
100000
50000
medAnnSal
2016
2018
2020
2022
ReportDate
JobTitle
Department Head
Senior Instructor I
Assistant Professor
Associate Professor
Instructor
Office Manager
Professor
Senior Instructor II

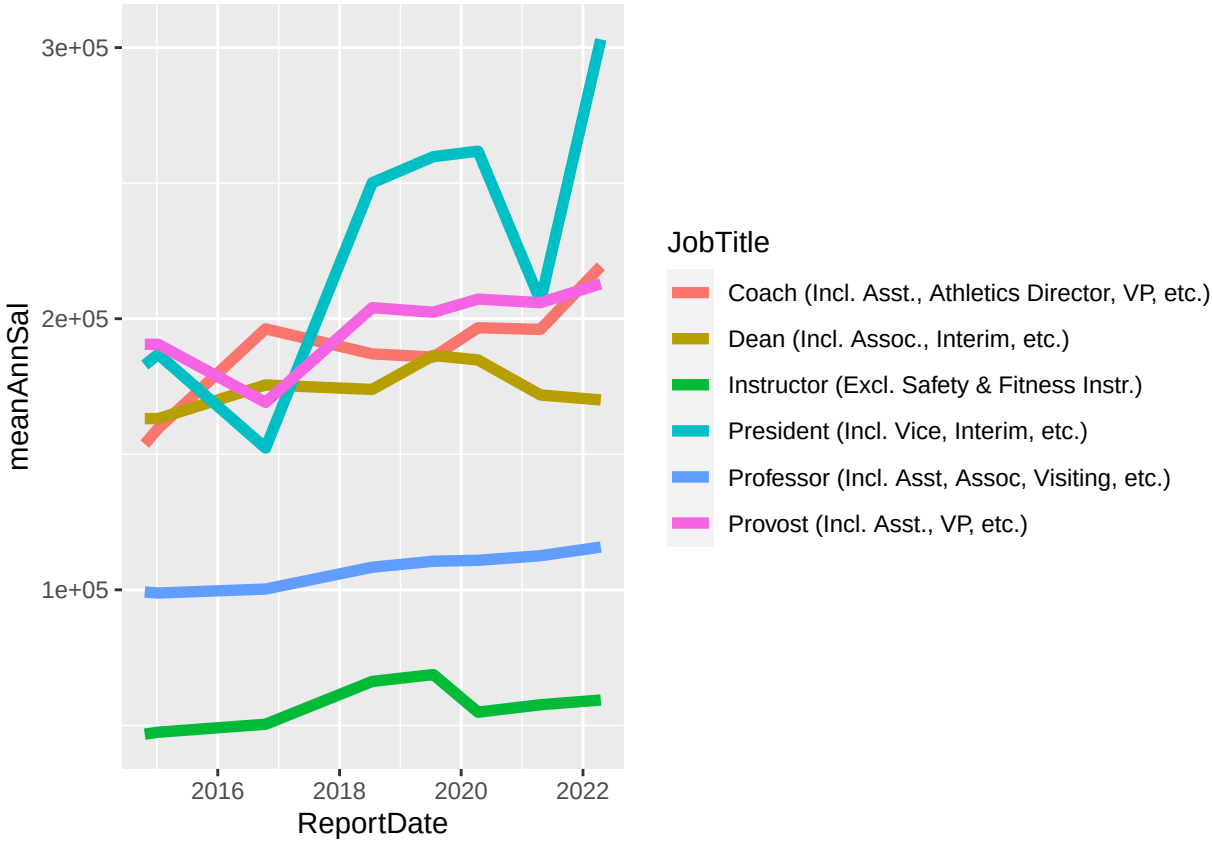
**Q2: How has Mean Salary Changed Overall?**

```
new.titles <- list(
    "Coaches"     = "Coach (Incl. Asst., Athletics Director, VP, etc.)",
    "Provosts"    = "Provost (Incl. Asst., VP, etc.)",
    "Deans"       = "Dean (Incl. Assoc., Interim, etc.)",
    "Presidents"  = "President (Incl. Vice, Interim, etc.)",
    "Professors"  = "Professor (Incl. Asst, Assoc, Visiting, etc.)",
    "Instructors" = "Instructor (Excl. Safety & Fitness Instr.)"
)

# still unclear if AnnSalary is adjusted for ApptPercent or not

employees.summary <- (
    employees.all
    %>% drop_na(AnnSalary)
    %>% filter(!str_detect(JobTitle, "Asst to"))
    %>% mutate(
            JobTitle = case_when(
                str_detect(JobTitle, "(Coach)|(Director)|(VP)") &
                    str_detect(JobOrgn, "YIA - Intercolleg Athletics")
                    ~ new.titles[["Coaches"]],
                str_detect(JobTitle, "Provost") ~ new.titles[["Provosts"]],
                str_detect(JobTitle, "Dean") ~ new.titles[["Deans"]],
                str_detect(JobTitle, "President") ~ new.titles[["Presidents"]],
                str_detect(JobTitle, "Professor") ~ new.titles[["Professors"]],
                str_detect(JobTitle, "Instructor") &
                    !str_detect(JobTitle, "(Motorcycle)|(Fitness)")
                    ~ new.titles[["Instructors"]],
                TRUE ~ ""
            )
        )
    %>% filter(JobTitle != "")
    %>% group_by(ReportDate, JobTitle)
        %>% summarize(meanAnnSal = mean(AnnSalary))
    %>% ungroup()
    %>% group_by(JobTitle)
        %>% mutate(
                meanAnnSal2014 = sum((year(ReportDate) == 2014)*(meanAnnSal)),
                PercDevFrom2014 = 100*(meanAnnSal-meanAnnSal2014)/meanAnnSal2014
            )
    %>% ungroup()
)

ggplot(employees.summary, aes(x=ReportDate, y=meanAnnSal, color=JobTitle)) +
        geom_line(lwd=2)
```


3e+05
2e+05
1e+05
meanAnnSal
2016
2018
2020
2022
ReportDate
JobTitle
Coach (Incl. Asst., Athletics Director, VP, etc.)
Dean (Incl. Assoc., Interim, etc.)
Instructor (Excl. Safety & Fitness Instr.)
President (Incl. Vice, Interim, etc.)
Professor (Incl. Asst, Assoc, Visiting, etc.)
Provost (Incl. Asst., VP, etc.)