Idea 2

Sonya Eason

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
```

```
library(patchwork)
library(lmerTest)
```

```
Loading required package: lme4
Loading required package: Matrix

Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

    expand, pack, unpack

Warning in check_dep_version(): ABI version mismatch:
lme4 was built with Matrix ABI version 2
Current Matrix ABI version is 1
Please re-install lme4 from source or restore original 'Matrix' package
```

```
Attaching package: 'lmerTest'

The following object is masked from 'package:lme4':

    lmer

The following object is masked from 'package:stats':

    step
```

```r
library(knitr)
library(broom)
```

```r
unemployment <- read_csv("data/Unemployment.csv")
```

```
Rows: 1848 Columns: 5
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (2): quarter, state
dbl (3): year, month, unemployment_rate

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
price <- read_csv("data/Price and Availability Data.csv")
```

```
Rows: 1680 Columns: 18
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (3): quarter, market, internal_class
dbl (15): year, RBA, available_space, availability_proportion, internal_clas...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
occupancy <- read_csv("data/Major Market Occupancy Data-revised.csv")
```

```
Rows: 190 Columns: 6
-- Column specification --------------------------------------------------------
```

```
Delimiter: ","
chr (2): quarter, market
dbl (4): year, ending_occupancy_proportion, starting_occupancy_proportion, a...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
leases <- read_csv("data/Leases.csv")
```

```
Rows: 194685 Columns: 35
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (17): quarter, monthsigned, market, building_name, building_id, address,...
dbl (18): year, zip, leasedSF, costarID, RBA, available_space, availability_...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
income <- read_csv("data/ACSST5Y2023.S2411-Data.csv", skip = 1)
```

```
New names:
Rows: 29 Columns: 291
-- Column specification
-------------------------------------------------------- Delimiter: "," chr
(8): Geography, Geographic Area Name, Estimate!!Median earnings (dolla... dbl
(282): Estimate!!Median earnings (dollars)!!Civilian employed population... lgl
(1): ...291
i Use `spec()` to retrieve the full column specification for this data. i
Specify the column types or set `show_col_types = FALSE` to quiet this message.
* `` -> `...291`
```

#sublet less

sublet availability and square footage

Is there association between sublet and leasing SF?

```
leases <- leases |>
  mutate(sublet_percent = sublet_availability_proportion *100)
```

```r
leases$date <- paste(leases$year, leases$monthsigned, sep = "-")
```

```r
mod <- lmerTest::lmer(
  leasedSF ~ sublet_percent + (1|date),
  data = leases
)
```

```r
summary(mod)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: leasedSF ~ sublet_percent + (1 | date)
   Data: leases

REML criterion at convergence: 2793830

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.061 -0.240 -0.149 -0.033 62.176

Random effects:
 Groups   Name        Variance  Std.Dev.
 date     (Intercept)  73841642  8593
 Residual             547407529 23397
Number of obs: 121674, groups:  date, 74

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)      3988.15    1008.23     71.02   3.956 0.000179 ***
sublet_percent   2750.12      66.24 121655.00  41.515  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr)
sublt_prcnt -0.113
```

```r
lmm <- function(x){
  x*2750.12 + 3988.15
}
```

```r
df <- data.frame(date = leases$date, sublet_percent = leases$sublet_percent, predleasedSF = l

df <- df |>
  filter(!is.na(sublet_percent)) |>
  filter(!is.na(predleasedSF))
```

why is facet wrap low……..

make one diagram showing noise from lmm

```r
markets_internal <- leases |>
  filter(!is.na(market)) |>
  filter(!is.na(internal_class)) |>
  filter(!is.na(sublet_availability_proportion)) |>
  group_by(market, internal_class) |>
  summarize(
    city = first(market),
    count = n(),
    sublet_percent = first(sublet_availability_proportion)*100,
    avsqft = mean(leasedSF)
  )
```

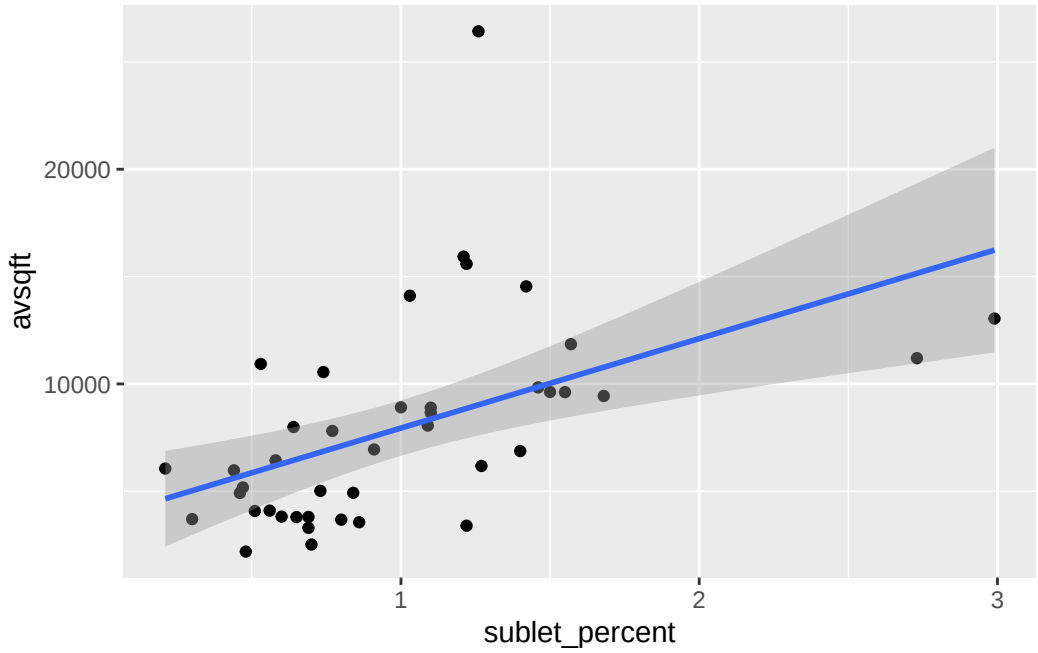
```
`summarise()` has grouped output by 'market'. You can override using the
`.groups` argument.
```

```r
markets_internal
```

```
# A tibble: 42 x 6
# Groups:   market [21]
   market    internal_class city      count sublet_percent avsqft
   <chr>     <chr>          <chr>     <int>          <dbl>  <dbl>
 1 Atlanta   A              Atlanta    3497           1.46  9836.
 2 Atlanta   O              Atlanta    2982           0.56  4097.
 3 Austin    A              Austin     1834           1.57 11850.
 4 Austin    O              Austin     2376           0.8   3676.
 5 Baltimore A              Baltimore  1401           0.64  7993.
 6 Baltimore O              Baltimore  1967           0.3   3701.
 7 Boston    A              Boston     2771           1.22 15591.
 8 Boston    O              Boston     3952           0.46  4918.
 9 Charlotte A              Charlotte  1238           0.74 10551.
10 Charlotte O              Charlotte  1585           0.69  3291.
# i 32 more rows
```

```
ggplot(markets_internal, aes(sublet_percent, avsqft)) +
  geom_point()+
  geom_smooth(method = "lm")
```

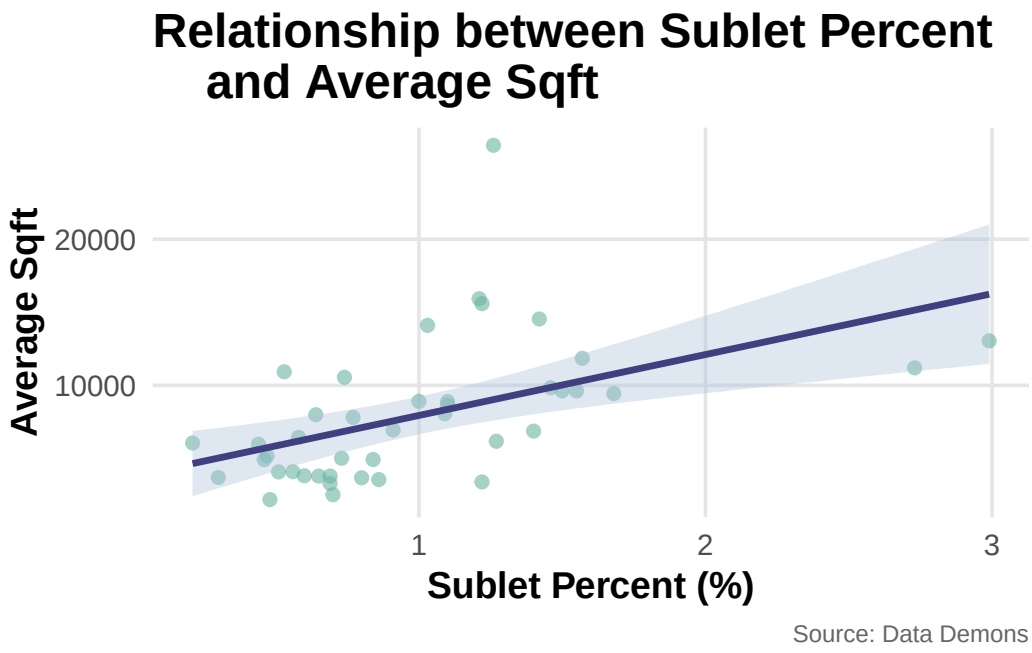
```
`geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(markets_internal, aes(sublet_percent, avsqft)) +
  geom_point(color = "#69b3a2", size = 2, alpha = 0.6) +  # soft teal points
  geom_smooth(method = "lm", color = "#404080", fill = "#b0c4de", se = TRUE, linetype = "soli
  theme_minimal(base_size = 14) +  # clean minimal theme with slightly larger font
  labs(
    title = "Relationship between Sublet Percent
    and Average Sqft",
    x = "Sublet Percent (%)",
    y = "Average Sqft",
    caption = "Source: Data Demons"
  ) +
  theme(
    plot.title = element_text(face = "bold", size = 18),
    plot.subtitle = element_text(size = 13, margin = margin(b = 10)),
    axis.title = element_text(face = "bold"),
```

```
    panel.grid.major = element_line(color = "gray90"),
    panel.grid.minor = element_blank(),
    plot.caption = element_text(size = 9, color = "gray40")
  )
```

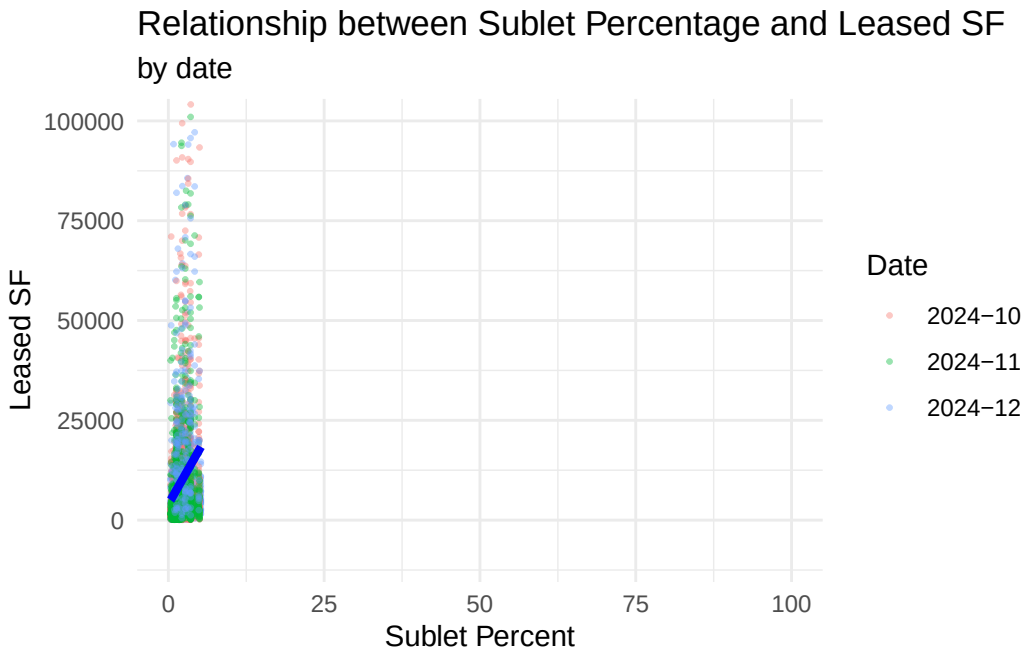
```
`geom_smooth()` using formula = 'y ~ x'
```




```
# Plot the data and the fitted regression lines
df |>
  filter(date%in%c("2024-12", "2024-11", "2024-10")) |>
  ggplot(aes(x = sublet_percent, y = leasedSF)) +
  geom_point(aes(color = factor(date)), alpha = 0.4, size = 0.5) +
  geom_line(aes(y = predleasedSF), color = "blue", size = 1.5)+
  #facet_wrap(~date)+
  theme_minimal() +
  coord_cartesian(ylim = c(-10000, 100000)) +
  scale_x_continuous(limits = c(0, 100)) + # or whatever range makes sense
# <-- adjust this range as needed
  labs(
    title = "Relationship between Sublet Percentage and Leased SF",
    subtitle = "by date",
```

```
    x = "Sublet Percent",
    y = "Leased SF",
    color = "Date"
  )
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



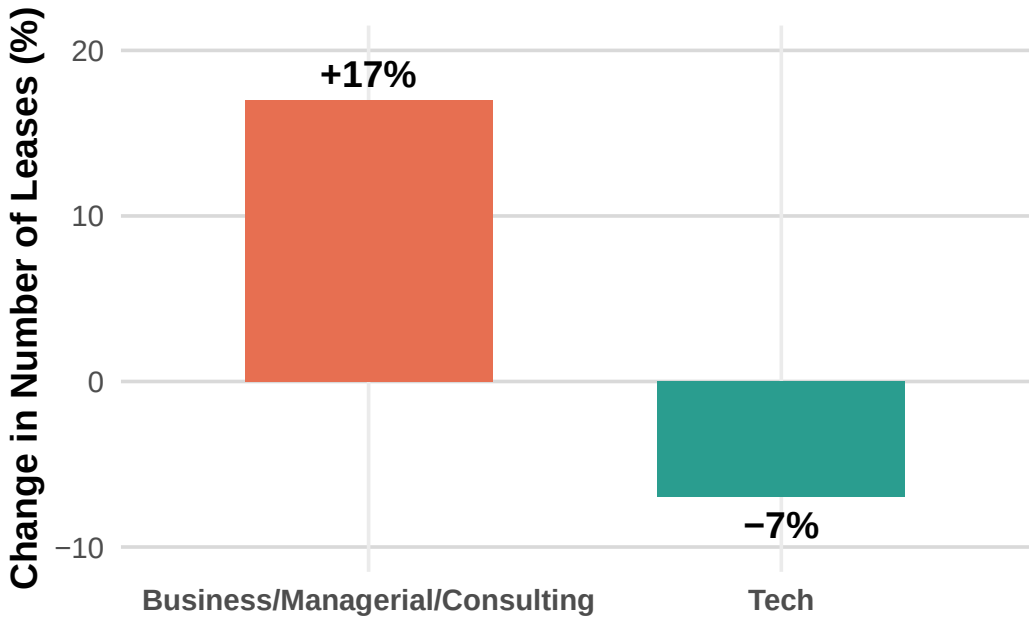
We use a mixture model, which serves to control for confounding variables that could potentially affect the relationship we're trying to understand. Here, this is useful because we are able to control for time.

```
library(ggplot2)
library(dplyr)

# Create the summary data
df <- data.frame(
  company_type = c("Tech", "Business/Managerial/Consulting"),
  percent_change_in_leases = c(-7, 17)
)

# Plot
```

```r
ggplot(df, aes(x = company_type, y = percent_change_in_leases, fill = company_type)) +
  geom_col(width = 0.6, show.legend = FALSE) +
  geom_text(aes(label = paste0(ifelse(percent_change_in_leases > 0, "+", ""), percent_change
            vjust = ifelse(df$percent_change_in_leases > 0, -0.5, 1.5),
            size = 5,
            fontface = "bold") +
   scale_fill_manual(values = c(
    "Tech" = "#2a9d8f",                                  # Soft teal
    "Business/Managerial/Consulting" = "#e76f51"     # Warm coral
  )) +
  labs(
    x = "",
    y = "Change in Number of Leases (%)"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", size = 18),
    plot.subtitle = element_text(size = 13, margin = margin(b = 10)),
    axis.text.x = element_text(face = "bold"),
    axis.title.y = element_text(face = "bold"),
    panel.grid.major.y = element_line(color = "gray85"),
    panel.grid.minor = element_blank()
  ) +
  ylim(-10, 20)
```

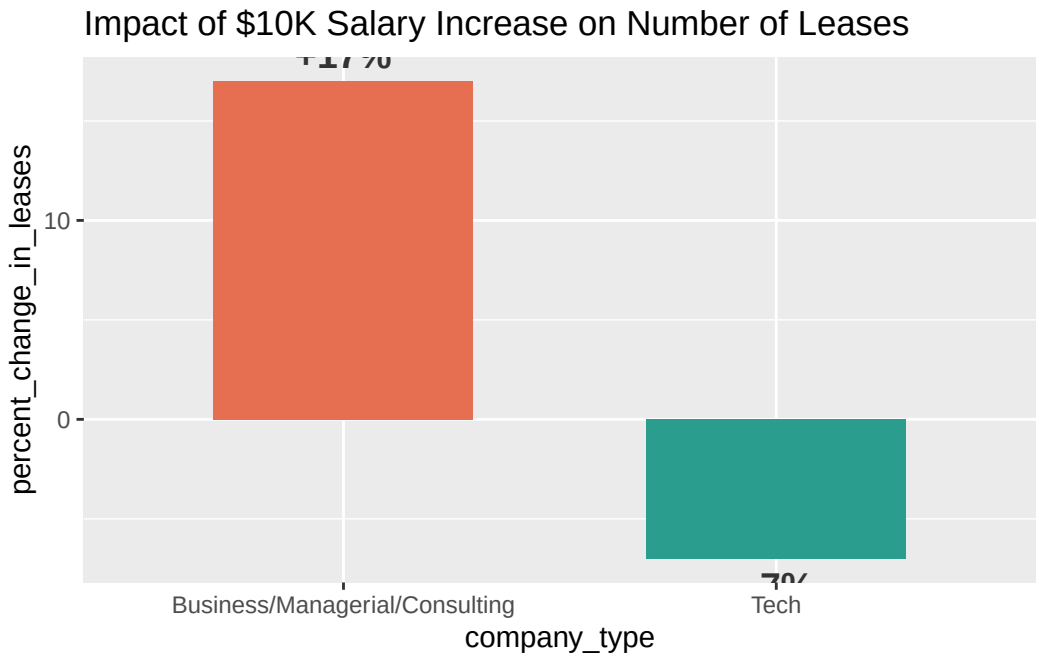



```r
library(ggplot2)

# Data
df <- data.frame(
  company_type = c("Tech", "Business/Managerial/Consulting"),
  percent_change_in_leases = c(-7, 17)
)

# Plot
ggplot(df, aes(x = company_type, y = percent_change_in_leases, fill = company_type)) +
  geom_col(width = 0.6, show.legend = FALSE) +
  geom_text(
    aes(label = paste0(ifelse(percent_change_in_leases > 0, "+", ""), percent_change_in_leas
    vjust = ifelse(df$percent_change_in_leases > 0, -0.5, 1.5),
    size = 5,
    fontface = "bold",
    color = "gray20"
  ) +
  scale_fill_manual(values = c(
    "Tech" = "#2a9d8f",                              # Soft teal
    "Business/Managerial/Consulting" = "#e76f51"    # Warm coral
  )) +
  labs(
```

```
    title = "Impact of $10K Salary Increase on Number of Leases",
)
```



make a visualization } –

```
leases |>
  filter(!is.na(leasedSF)) |>
  summarize(av = mean(leasedSF))
```

```
# A tibble: 1 x 1
     av
  <dbl>
1 7419.
```

```
leases |>
  group_by(sublet_availability_proportion) |>
  count()
```

```
# A tibble: 372 x 2
# Groups:   sublet_availability_proportion [372]
   sublet_availability_proportion     n
```

```
                            <dbl> <int>
 1                         0.0017    58
 2                         0.0021    60
 3                         0.0023    64
 4                         0.0025   145
 5                         0.0029   152
 6                         0.003     97
 7                         0.0032    58
 8                         0.0033   286
 9                         0.0034   194
10                         0.0035   340
# i 362 more rows
```