

Permanencia Pragma

Análisis exploratorio

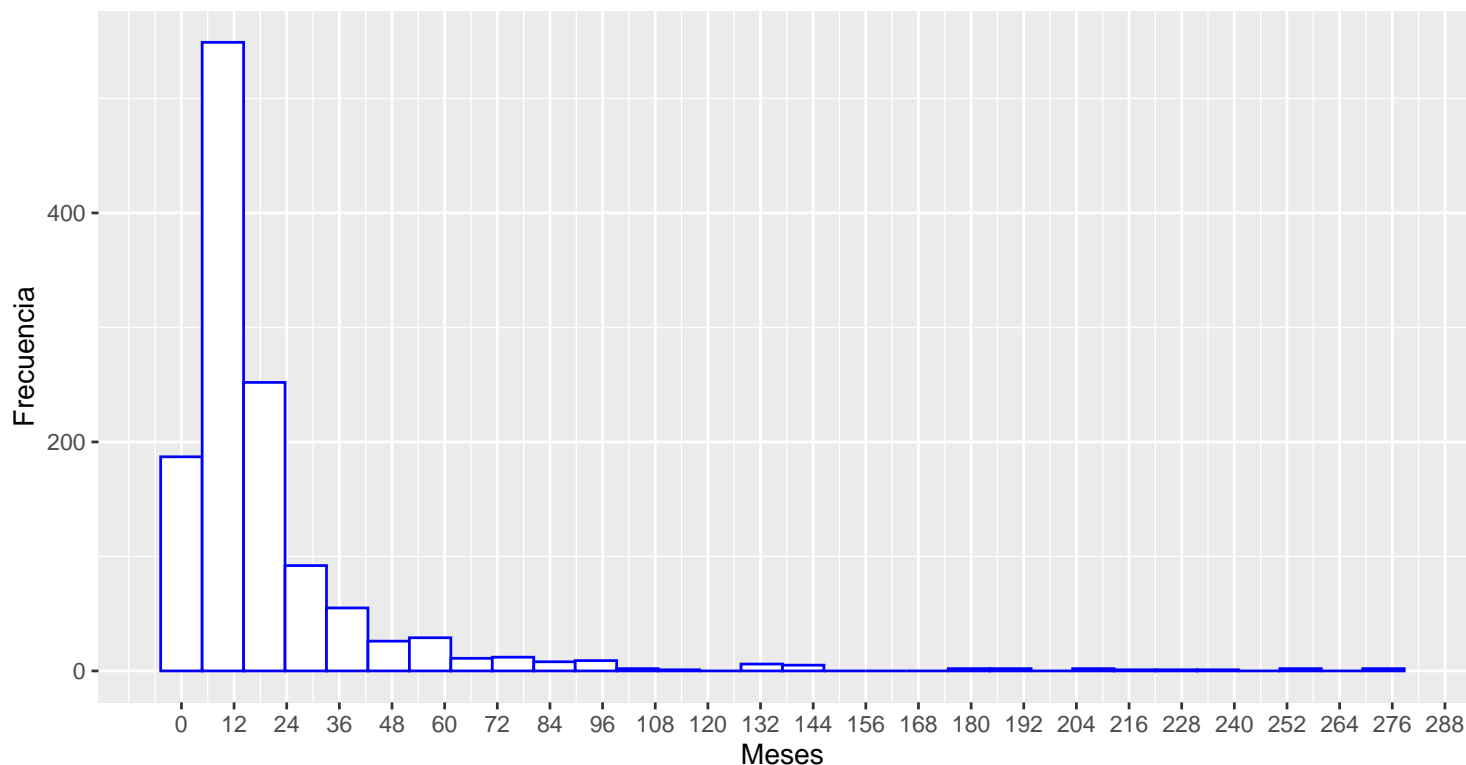
Vistazo a los datos

```
'data.frame':  1329 obs. of  9 variables:
 $ id      : chr  "005428920349380a73fff626ecfac47c" "005c436c2cd0d824eac8fc7047198308" "006701a9d614
 $ area    : chr  "CUSTOMER SUCCESS" "CUSTOMER SUCCESS" "CUSTOMER SUCCESS" "CUSTOMER SUCCESS" ...
 $ sex     : chr  "FEMENINO" "FEMENINO" "MASCULINO" "MASCULINO" ...
 $ age     : num  27 43 24 25 31 36 44 23 22 28 ...
 $ antiquity: num  9 57 0 21 93 35 32 0 5 10 ...
 $ salary  : int  4000000 33000000 2050000 5500000 3406000 8500000 3006000 3000000 2320000 3006000 ..
 $ state   : chr  "active" "active" "retired" "active" ...
 $ chapter : chr  "EXPERIENCIA_USUARIO" "HEAD" "BACKEND" "BACKEND" ...
 $ seniority: chr  "ADVANCED" "MASTER" "TRAINEE" "ADVANCED" ...
```

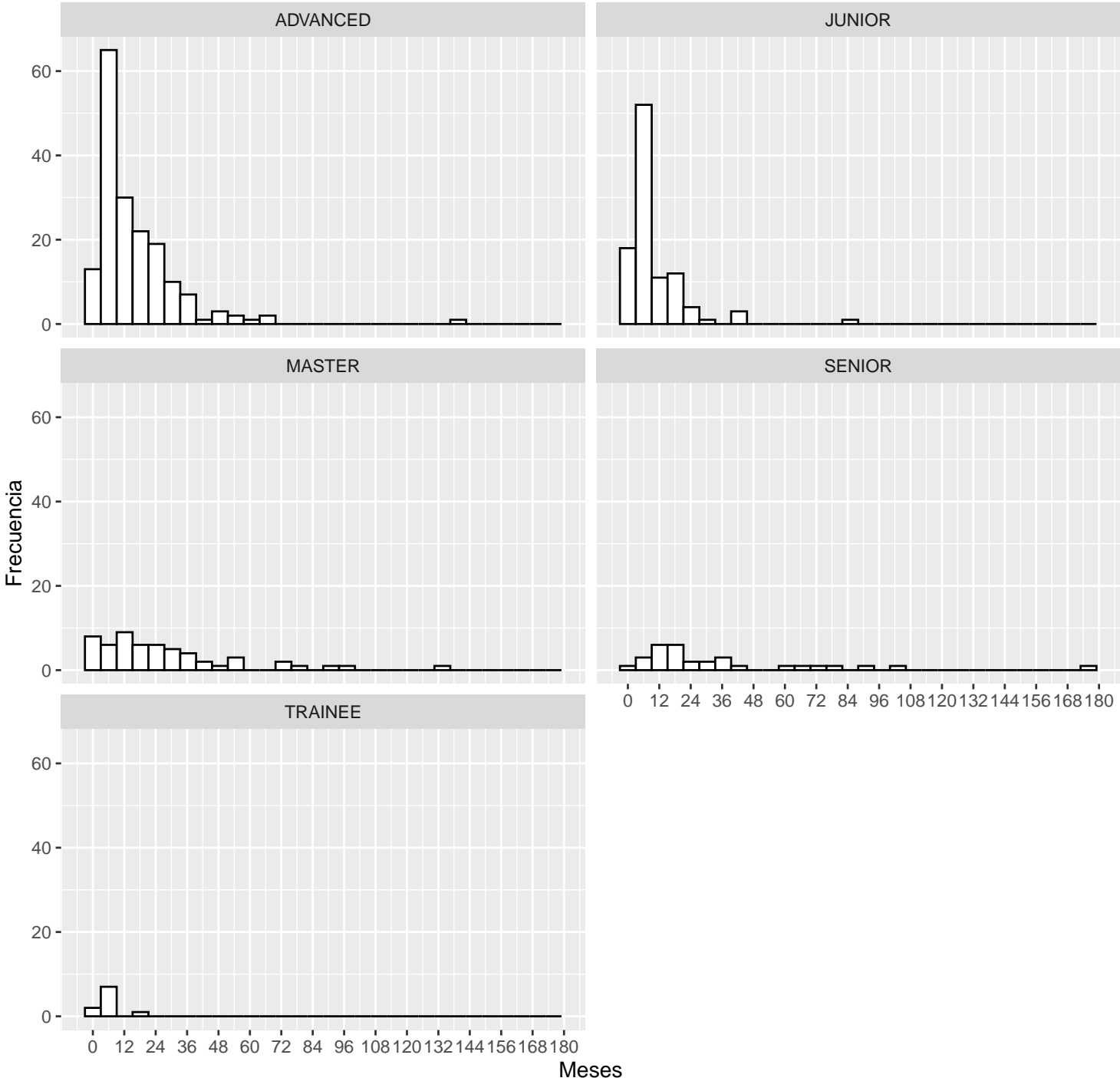
No se consideraron los 72 salarios menores a 100 mil.

Distribución de los tiempos de permanencia

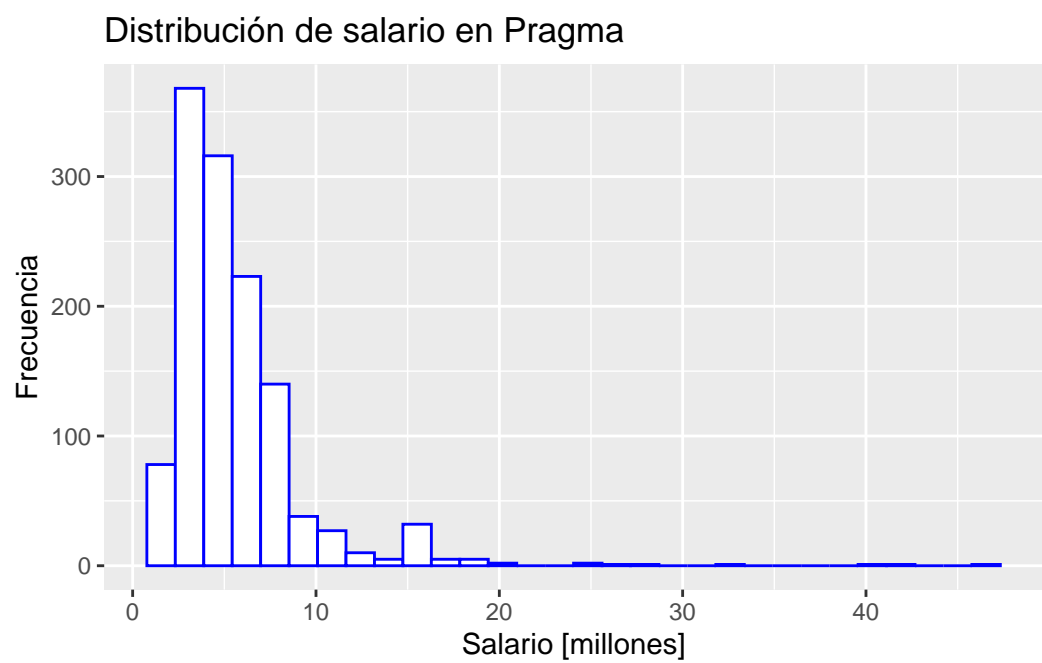
Tiempo de permanencia en Pragma (todos los empleados)



Empleados retirados – ¿Cuándo se van?



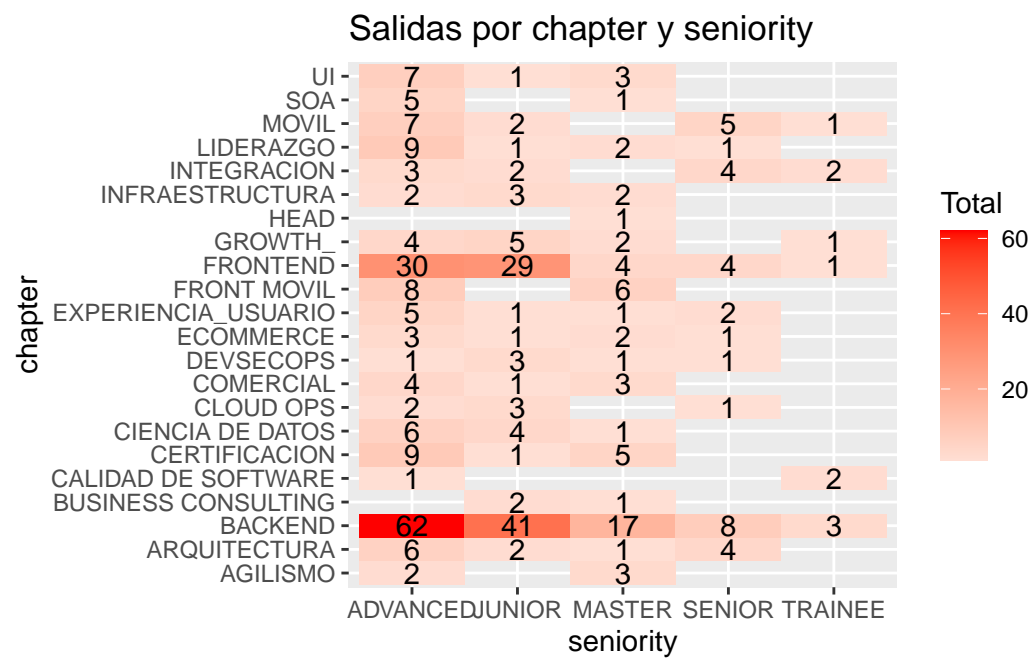
Distribución de los salarios



A partir del histograma anterior, se deciden las siguientes clasificaciones para el salario (en millones):

- nivel1: De 0 a 2 millones.
- nivel2: De 2 a 4 millones.
- nivel3: De 4 a 7 millones.
- nivel4: De 7 a 10 millones.
- nivel5: Más de 10 millones.

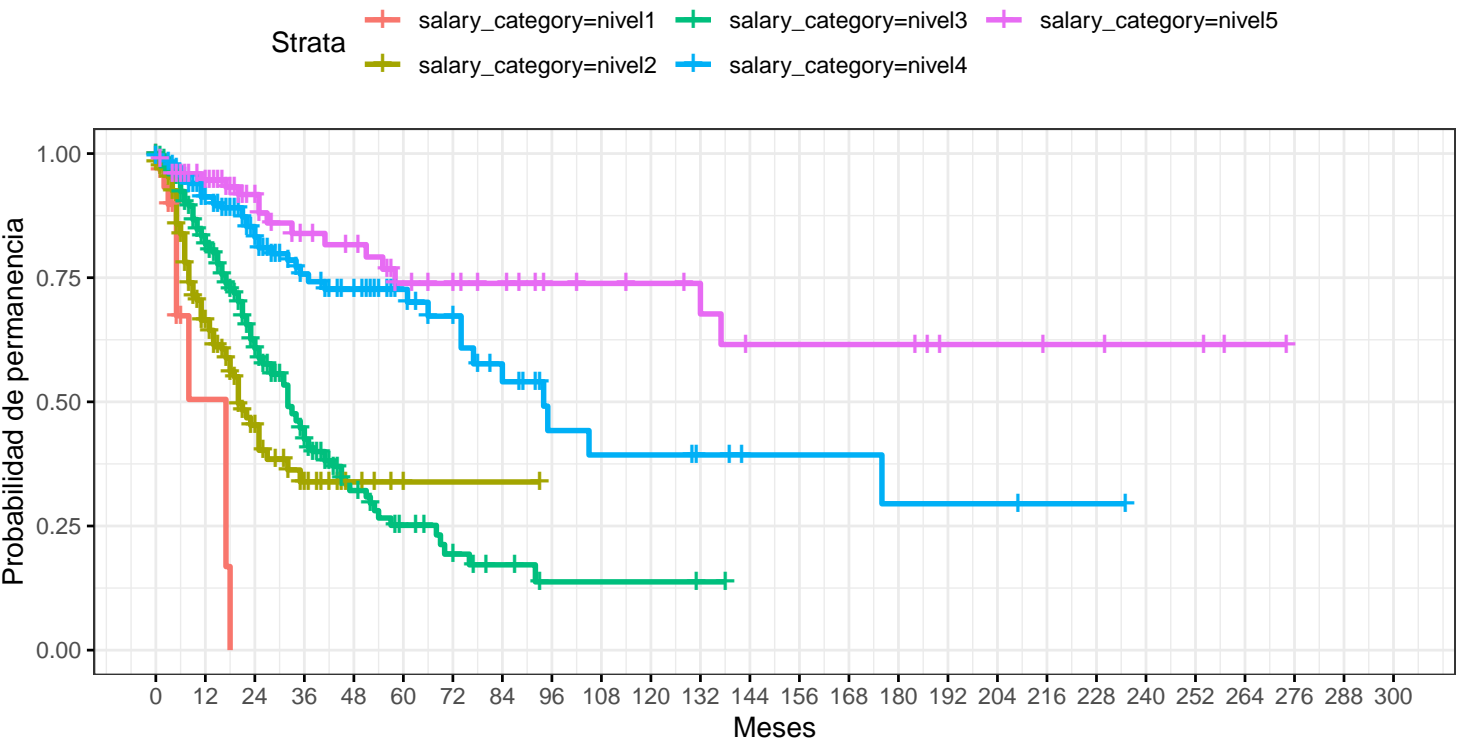
Mapa de calor



Curvas de supervivencia

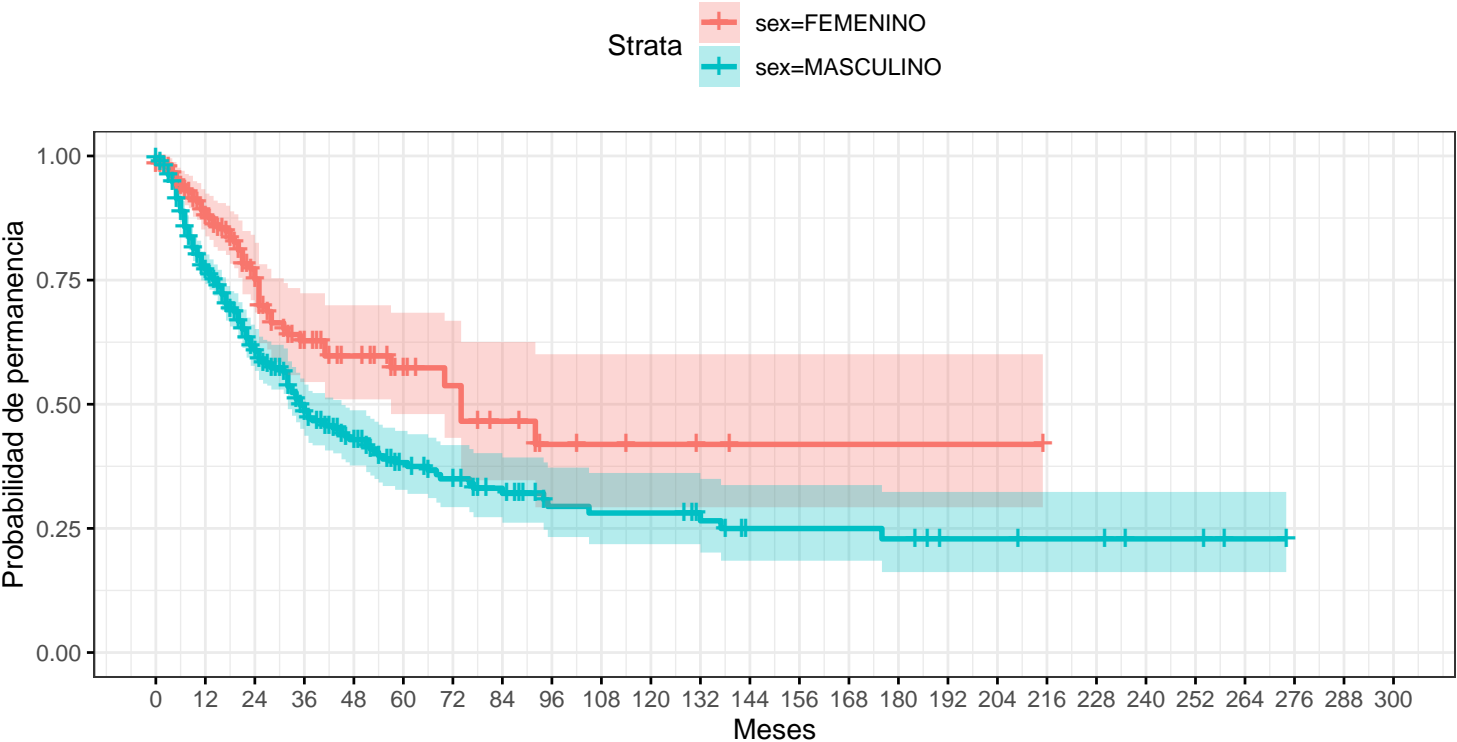
Categoría Salario

Curva de supervivencia – salary_category



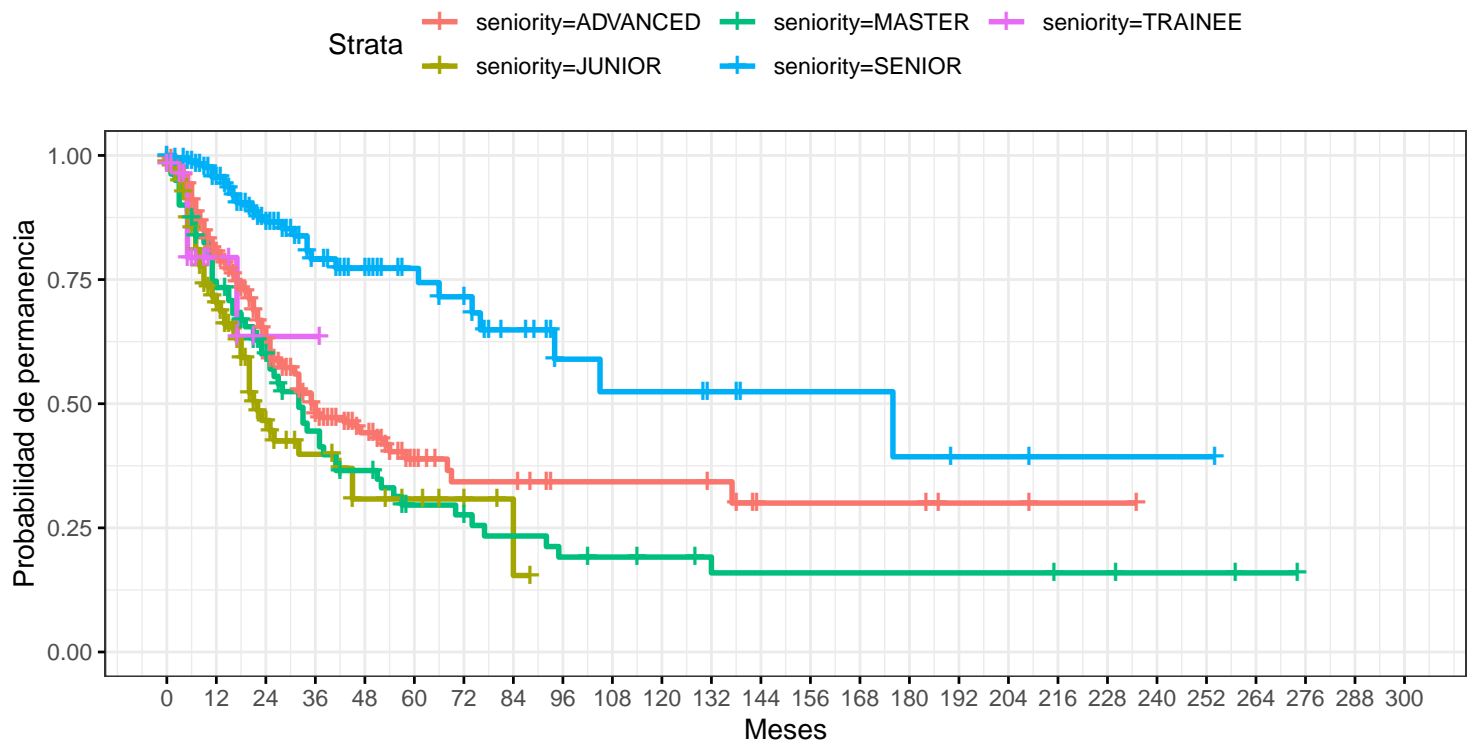
Sexo

Curva de supervivencia – sex



Seniority

Curva de supervivencia – seniority



Modelo de Regresión Logística

```
empleados2$seniority <- relevel(factor(empleados2$seniority), ref = "TRAINEE")

reg_logistica <- glm(status ~ salary + sex + seniority,
  data = empleados2,
  family = "binomial")

summary(reg_logistica)
```

Call:

```
glm(formula = status ~ salary + sex + seniority, family = "binomial",
  data = empleados2)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0013	-0.8607	-0.6978	1.3153	2.4184

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.51256	0.38395	-3.939	0.00008166782447 ***
salary	-0.17560	0.03088	-5.687	0.00000001296569 ***
sexMASCULINO	0.44418	0.16979	2.616	0.008895 **
seniorityADVANCED	1.27750	0.37320	3.423	0.000619 ***
seniorityJUNIOR	0.79288	0.37110	2.137	0.032631 *

```

seniorityMASTER    3.79150    0.50498    7.508 0.000000000000006 ***
senioritySENIOR    0.97406    0.43389    2.245    0.024770 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 1532.1  on 1256  degrees of freedom
Residual deviance: 1400.7  on 1250  degrees of freedom
AIC: 1414.7

```

Number of Fisher Scoring iterations: 5

INTERPRETACIONES

salary: Por cada unidad adicional de salario (1 M) las chances de quedarse aumentan un $1/\exp(-0.15261) - 1 = 16.4\%$.

sex: Los hombres tienen un $\exp(0.48014) - 1 = 61.6\%$ más de chances de salir si se comparan con una mujer de las mismas condiciones (seniority, salary).

seniority: En orden de chances de salida: MASTER > ADVANCED > SENIOR > JUNIOR > TRAINEE (por lo coeficientes, el nivel de referencia lo cambié a TRAINEE). La interpretación puede ser del estilo: un ADVANCED tiene más chances de salir que un JUNIOR de las mismas condiciones (sex, salary), sin embargo, note que el salario por lo general va a ser muy diferente. Podría existir un problema de multicolinealidad entre salary y seniority ya que están altamente relacionados (con más seniority se gana más.). Sugiero tener cuidado en esta interpretación e insistir más en la parte exploratoria (para dar insights).

Recomendaciones

Variables a incluir:

- KC
- Tiempo desde el último aumento de salario
- Rol
- Nivel académico

Se sugiere que en una segunda iteración del proyecto, se obtengan datos históricos por empleados. Con los datos actuales, tenemos solo un “punto en el tiempo”. Por ejemplo, puede que cierta persona sea master y se ha quedado mucho tiempo en Pragma, pero los datos no indican si empezó como junior, advanced o cuántas promociones ha recibido.

Por otra parte, con la variables ‘rol’ y ‘nivel académico’ será posible clusterizar ciertos pragmáticos y relacionarlos con personas de un nivel académico, rol y experiencia similar; y determinar si tienen un salario inferior al de sus pares con características similares y si esto influye a la no permanencia.

A presentar a stakeholders (en análisis descriptivo)

Desde el análisis descriptivo que se vaya a presentar en el proyecto, considero que se puede aportar mucho valor, por el momento, sugiero incluir:

- Ratio de salidas por roles/o chapters/o KC. ¿cuáles son los roles/chapters con más salidas?

Modelación

Algunas ideas respecto

- Modelar el tiempo de permanencia (análisis de supervivencia o modelos lineales generalizados)