Xinavane data overview

Laia Cirera
Joe Brew
Elisa Sicuri

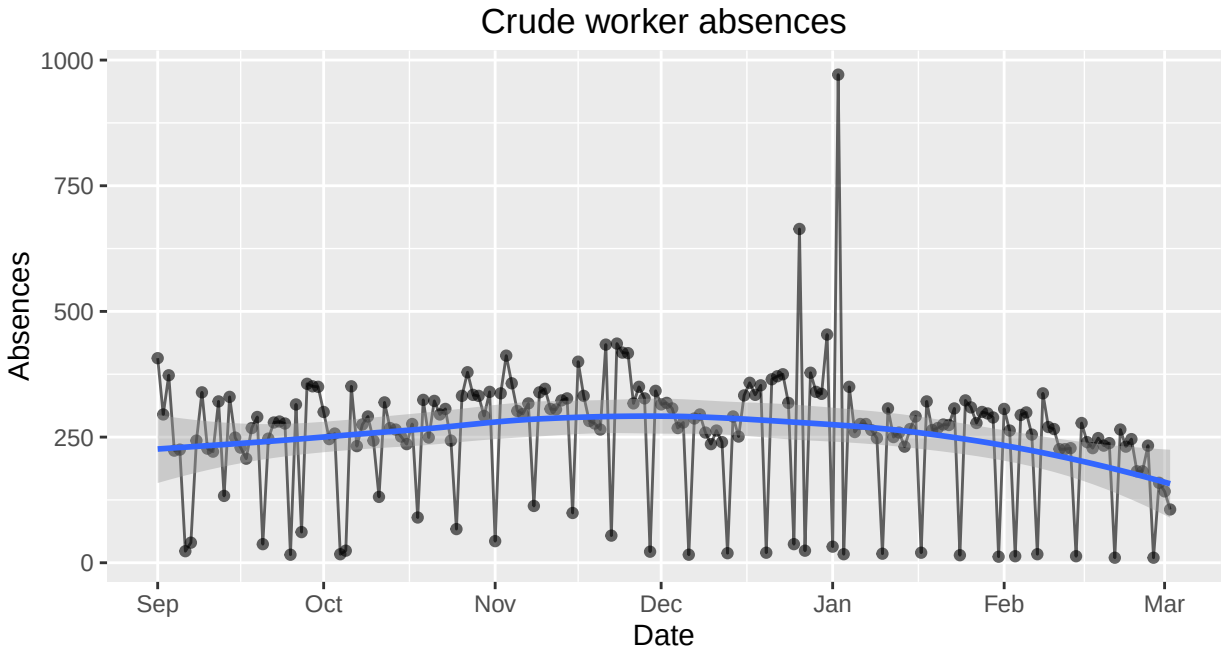
March 8, 2016

Contents

Exploration

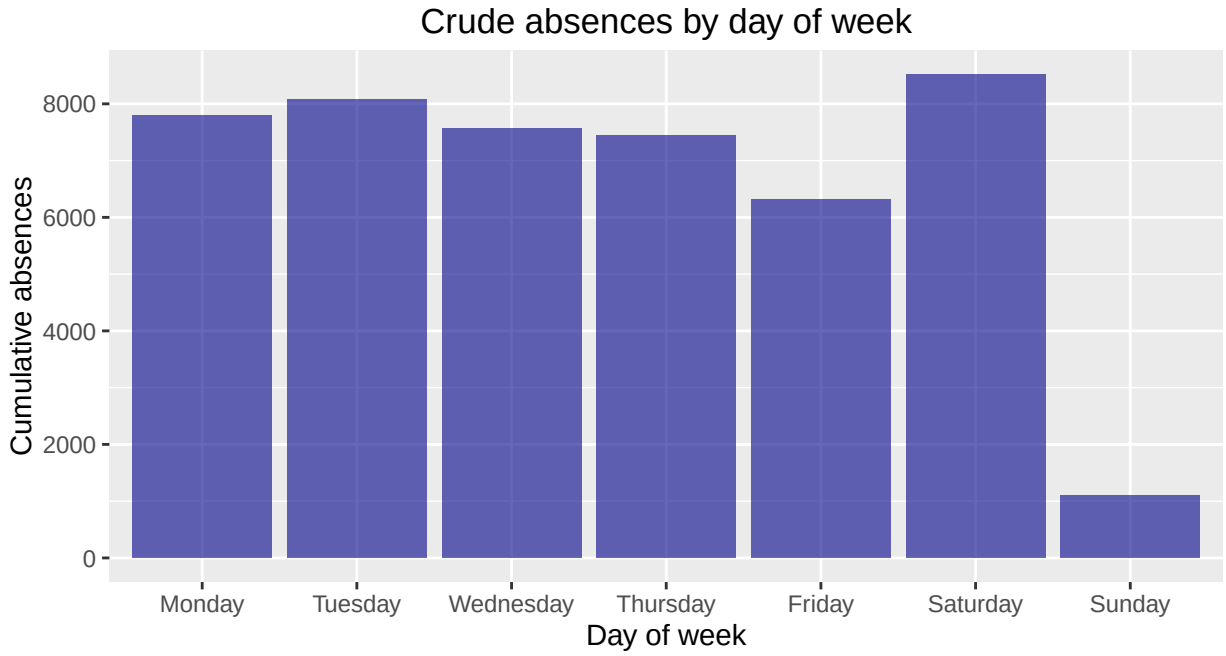
Absences

Absenteeism data spans from September 01, 2015 until March 02, 2016. On average, there are 253 absences per day. Though there appears to be some longer-term variation in the below chart, it's clear that the most important factor is weekly seasonality:

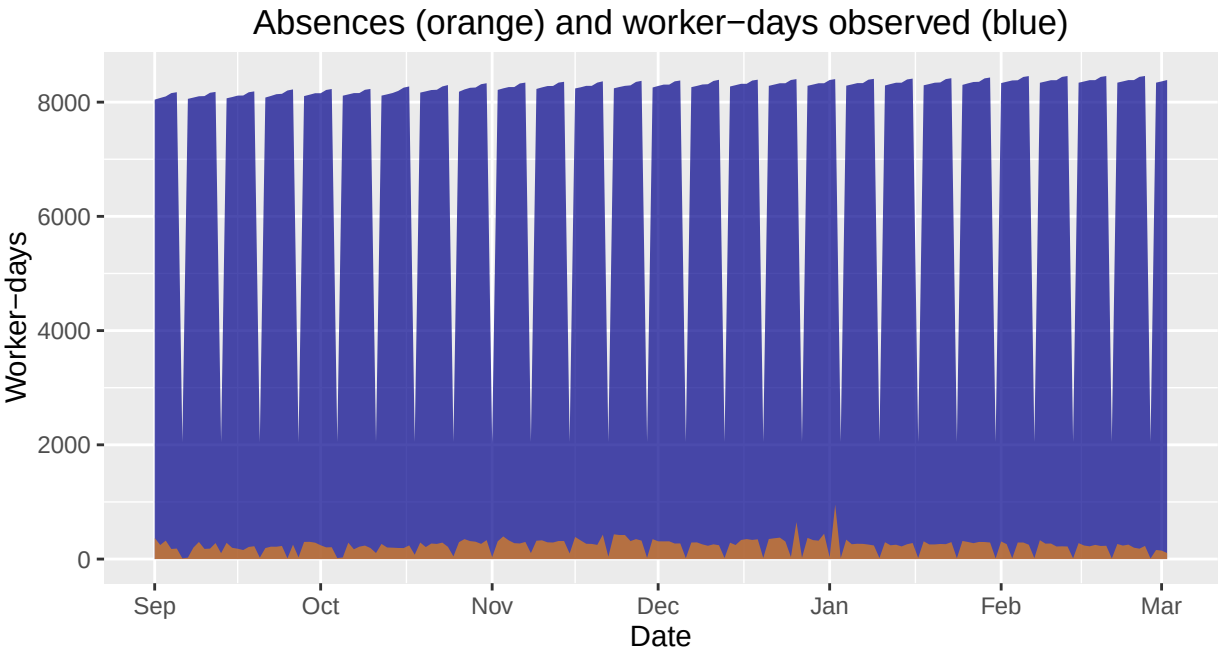


Adjustment for worker days

Crude absence data is relatively useless, given that it doesn't take into account the number of workers "susceptible" of absence on any given day (ie, the number of workers who were *supposed* to work). This explains why there are much fewer absences at certain times (Sunday) relative to others.



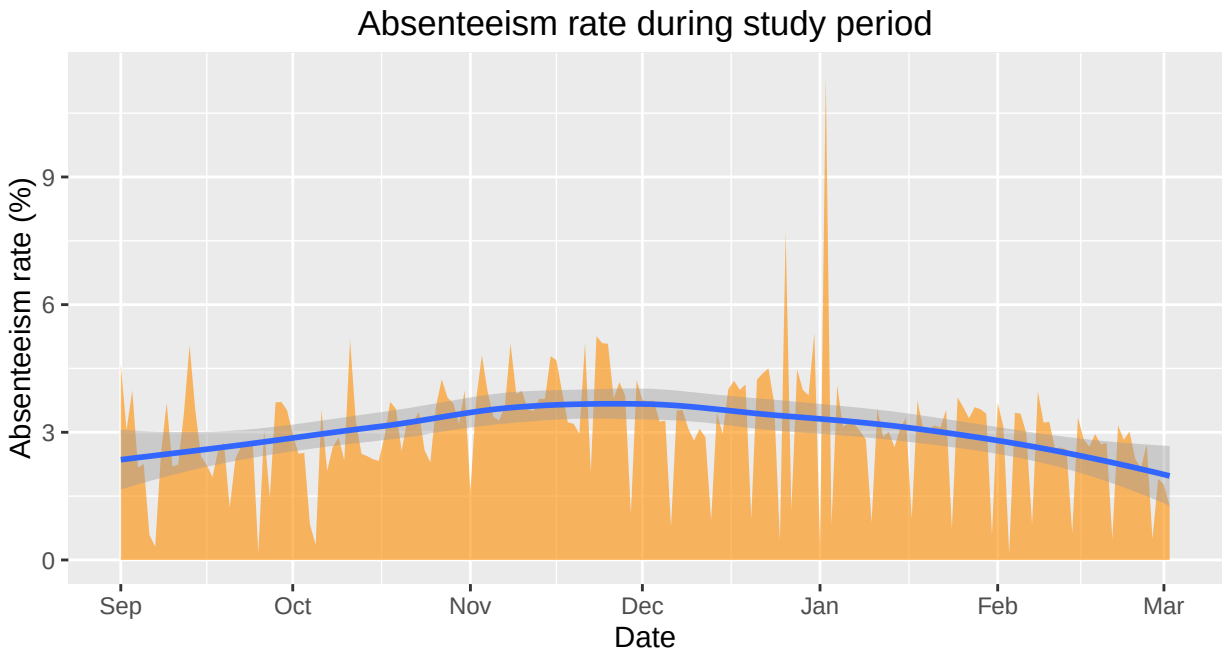
To account for the above issues, we can instead calculate an *absenteeism rate*, taking into account the employment beginning and end dates, leave statuses and working schedule of all the workers in the `Xinavane_plantilla trabajadores_agriculture_joe.xls` dataset. Essentially, we calculate the number of eligible workers for each day (the denominator in our rate). The result looks like this:



Absenteeism b

Absenteeism rate

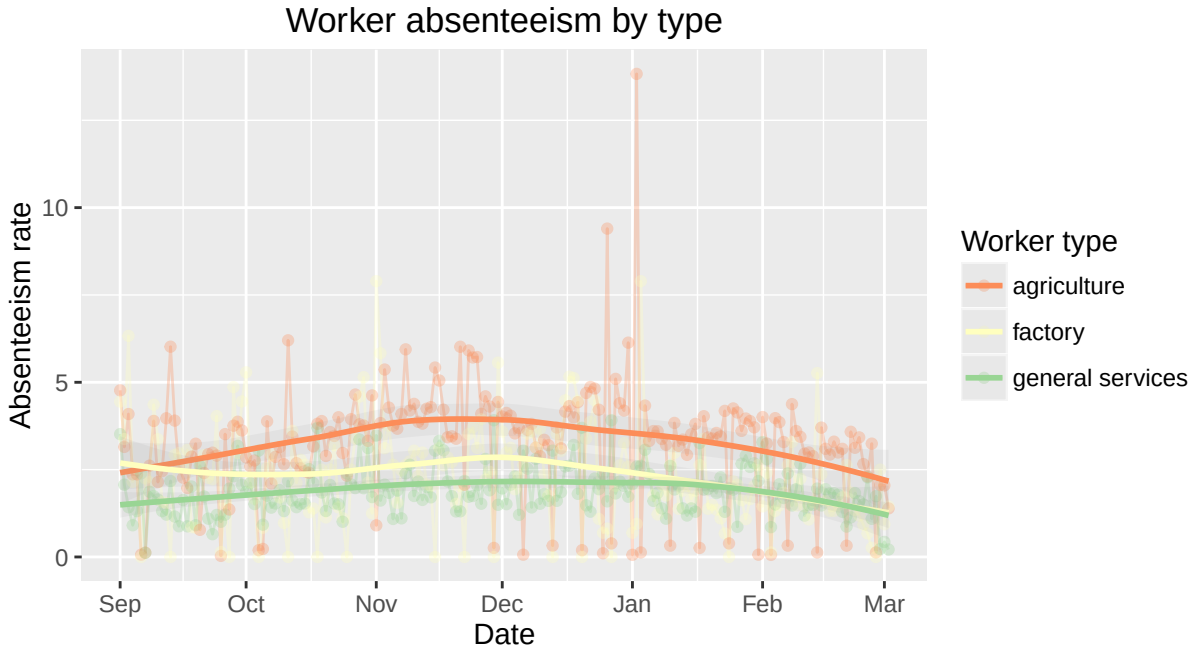
The absenteeism rate is calculated by dividing the number of worker-absences per day over the number of worker-days on the schedule (ie, those who were *supposed* to work that day). Over time, absenteeism at Xinavane appears like this.



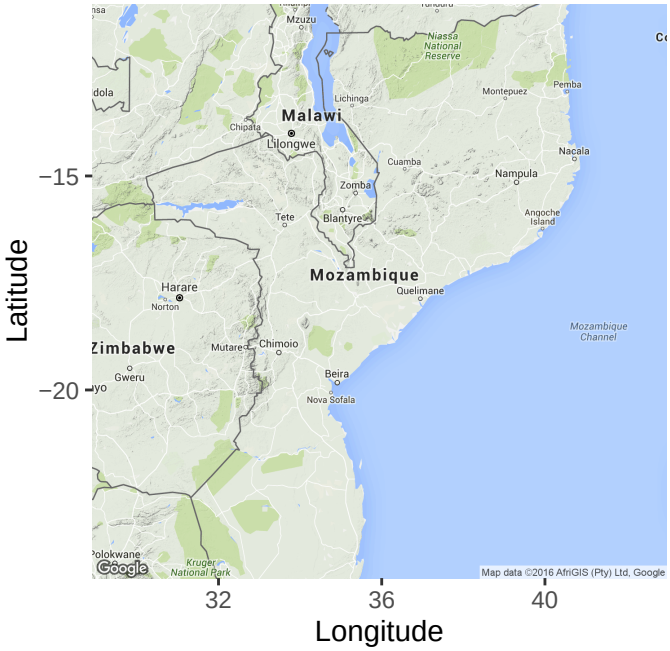
The daily average absenteeism rate (adjusted for worker eligibility) is 3.21%.

Absenteeism rate by worker type

Agricultural workers have significantly higher absenteeism than those that work in the factory or in general services.



Geography

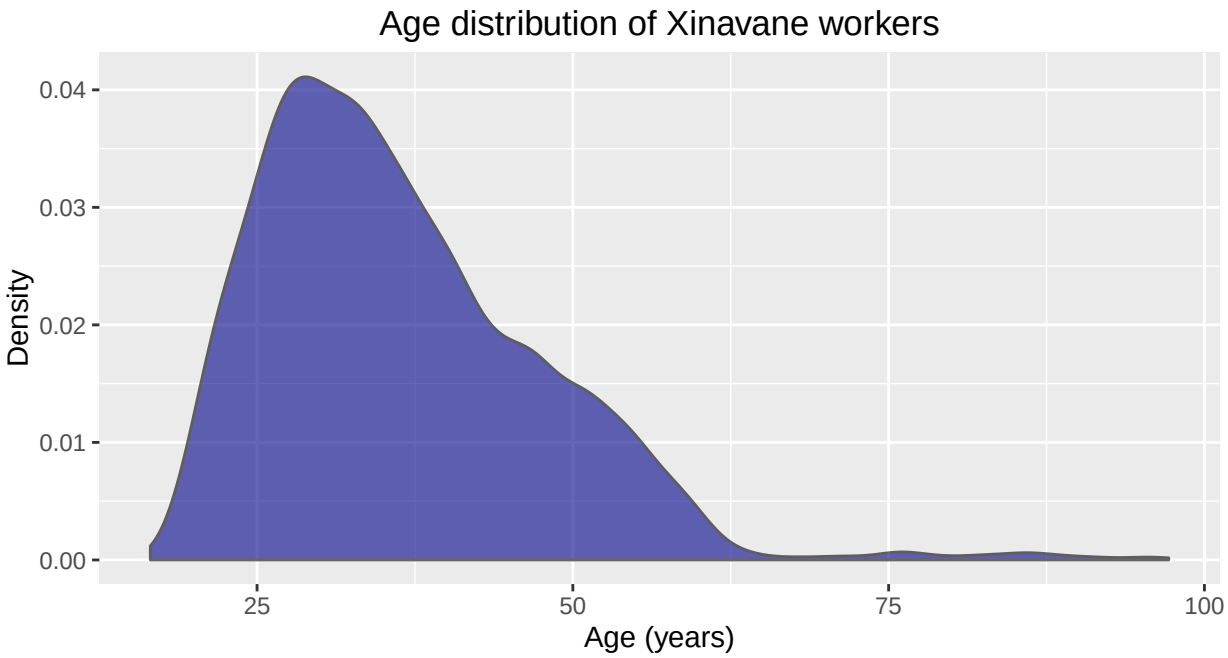


Unfortunately, the geographic data in the workers dataset is of inconsistent and low quality. The `home location` field contains many missing and proprietary names, and even when addresses are provided, modern geocoding API's are largely unable to match at a level more granular than the city/town.

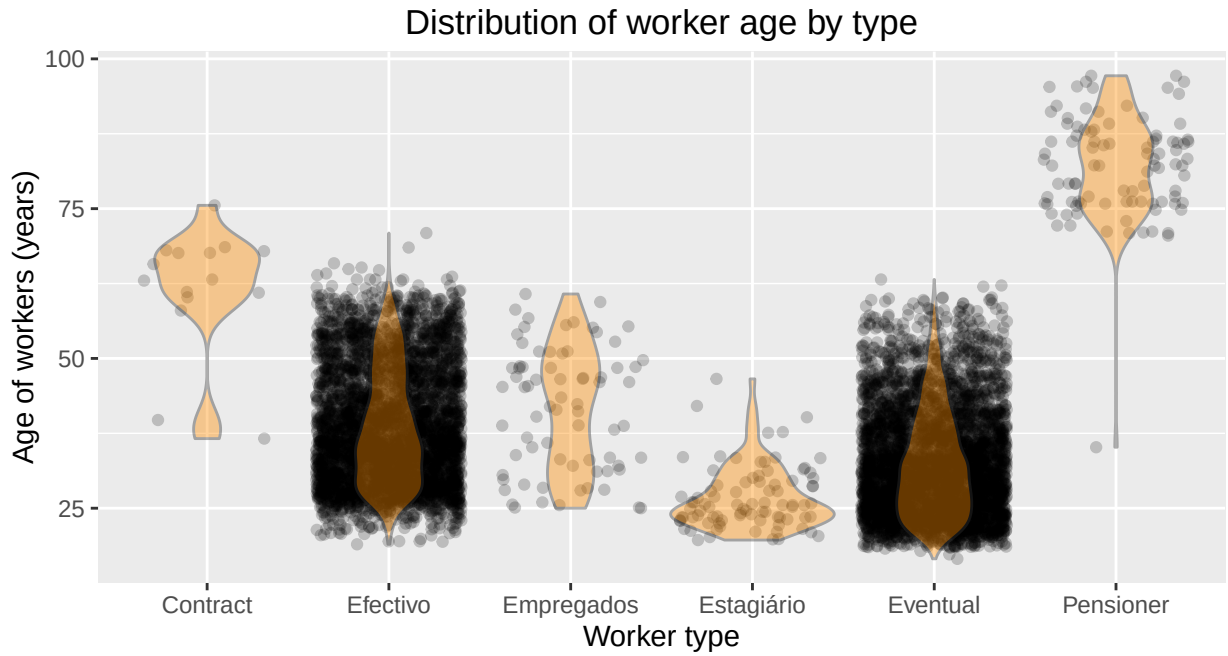
However, given the relatively small size of the number of workers (6,185), and the fact that a majority are on-site or nearby, manual geocoding of each location would be feasible (using typical hand-held cellular devices).

Age

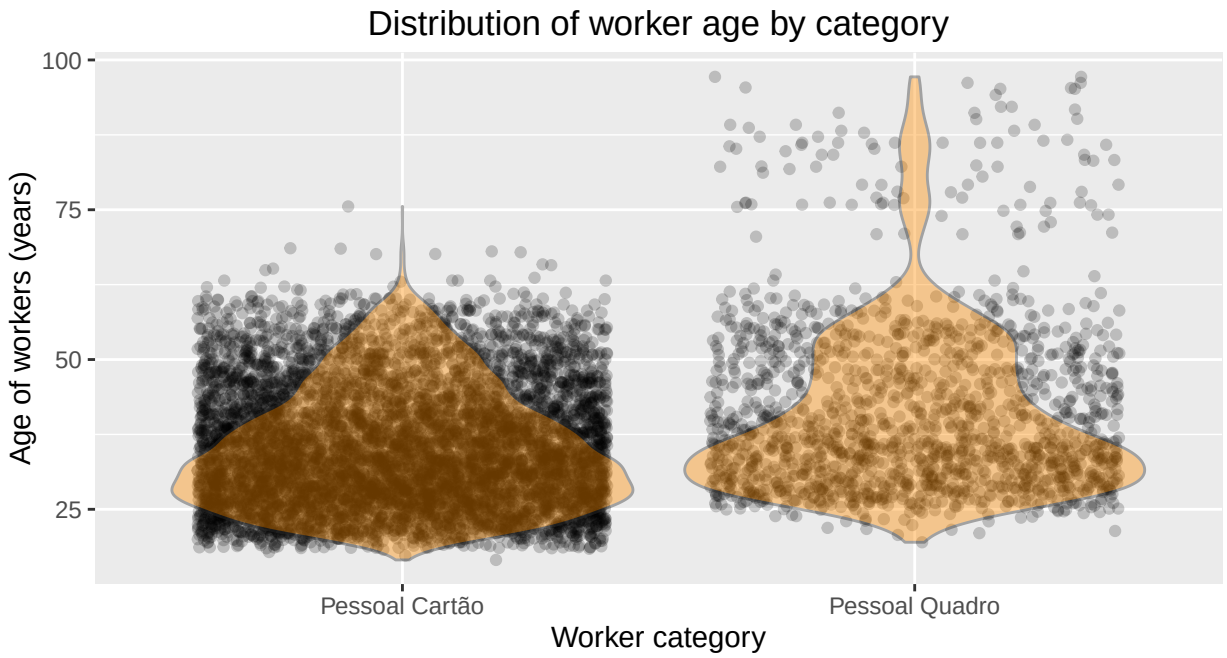
To the extent that absenteeism is confounded with health (which is in turn affected by age), it's important to note the age of workers.



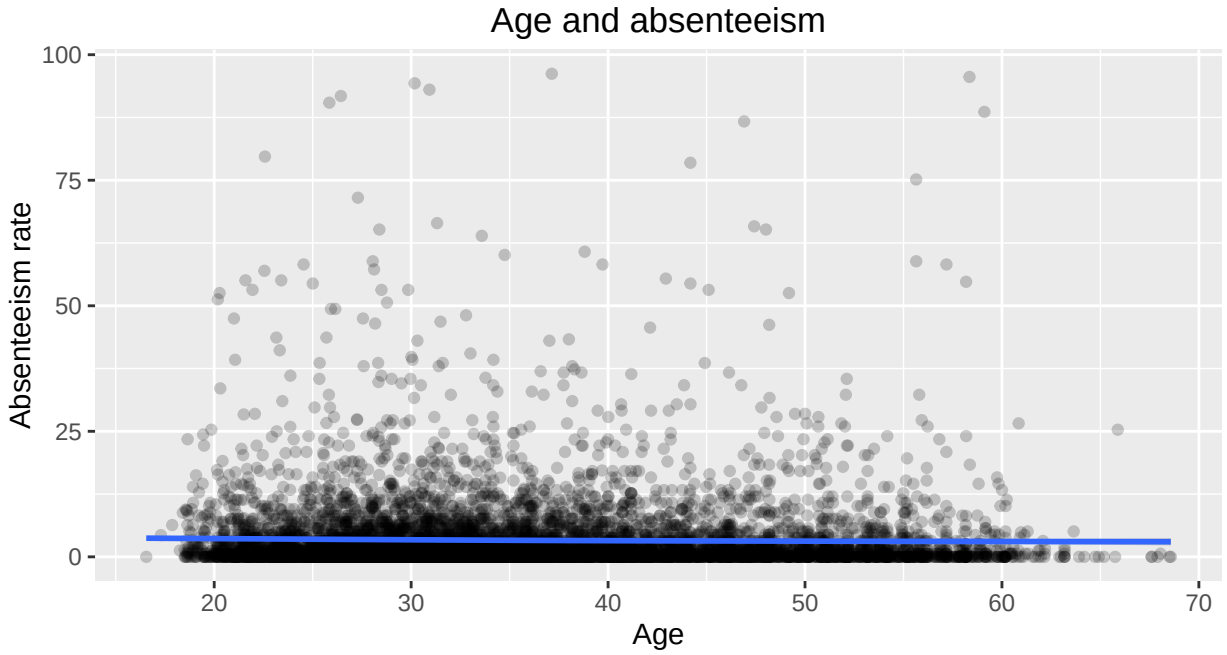
It's worthwhile to note that the age of workers varies significantly by *type*:



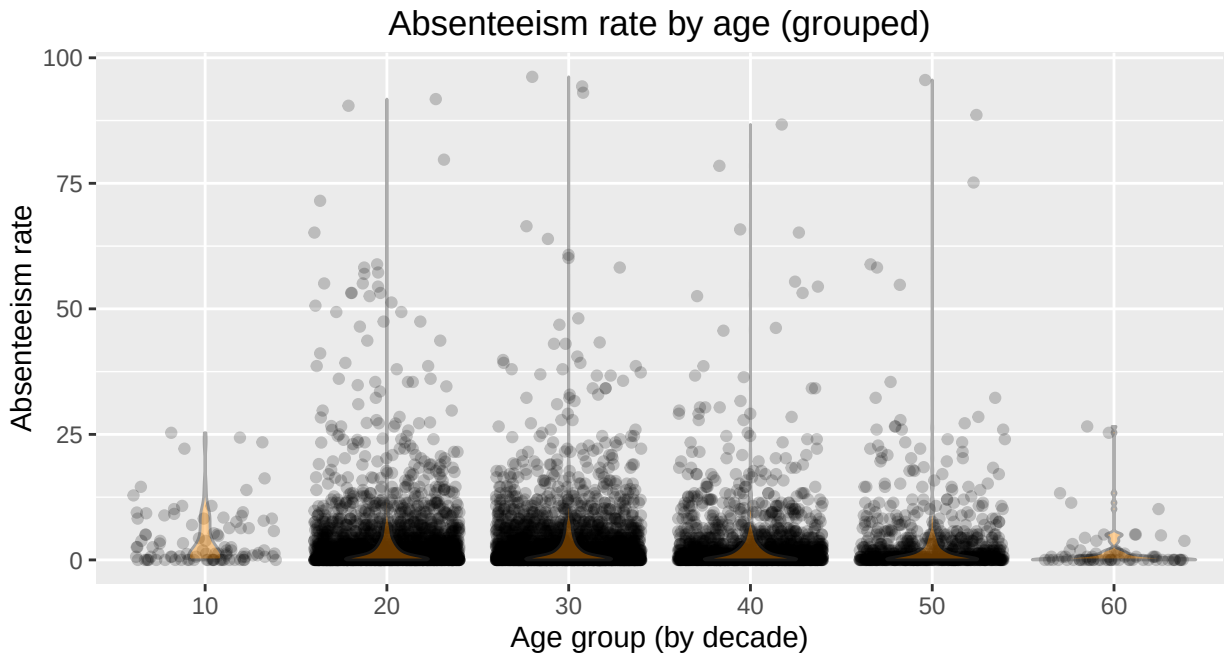
Worker age also varies significantly by category:



Age does not appear to be correlated with absences:

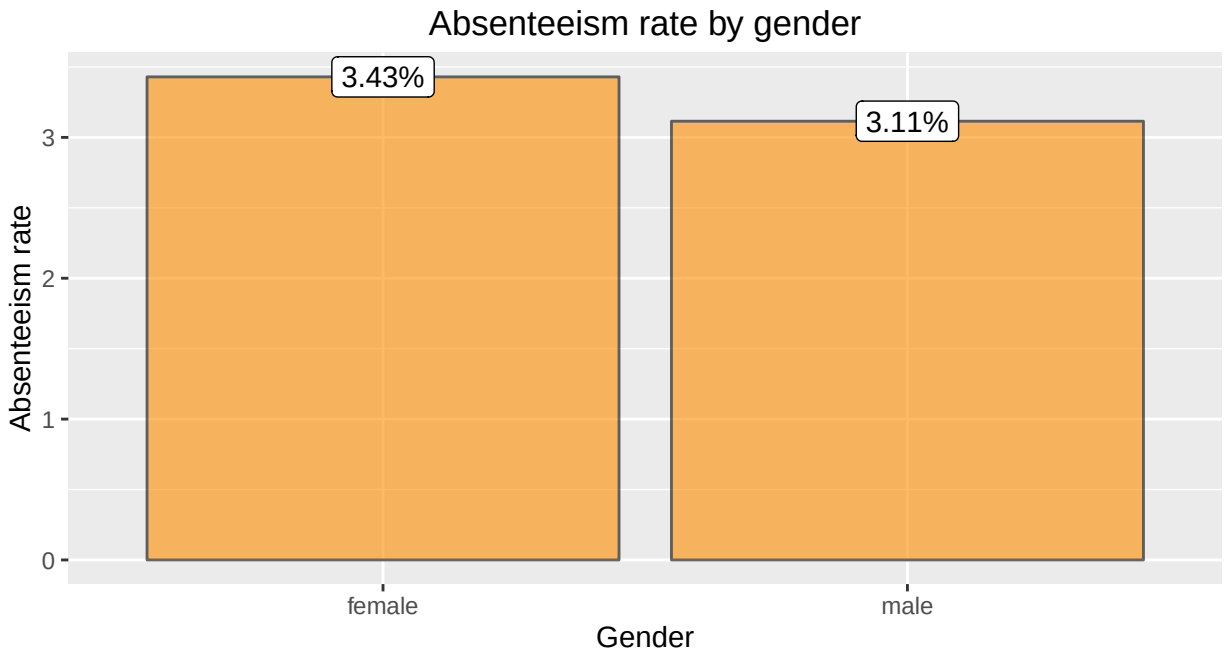


When we group by decade, a trend becomes slightly more apparent:

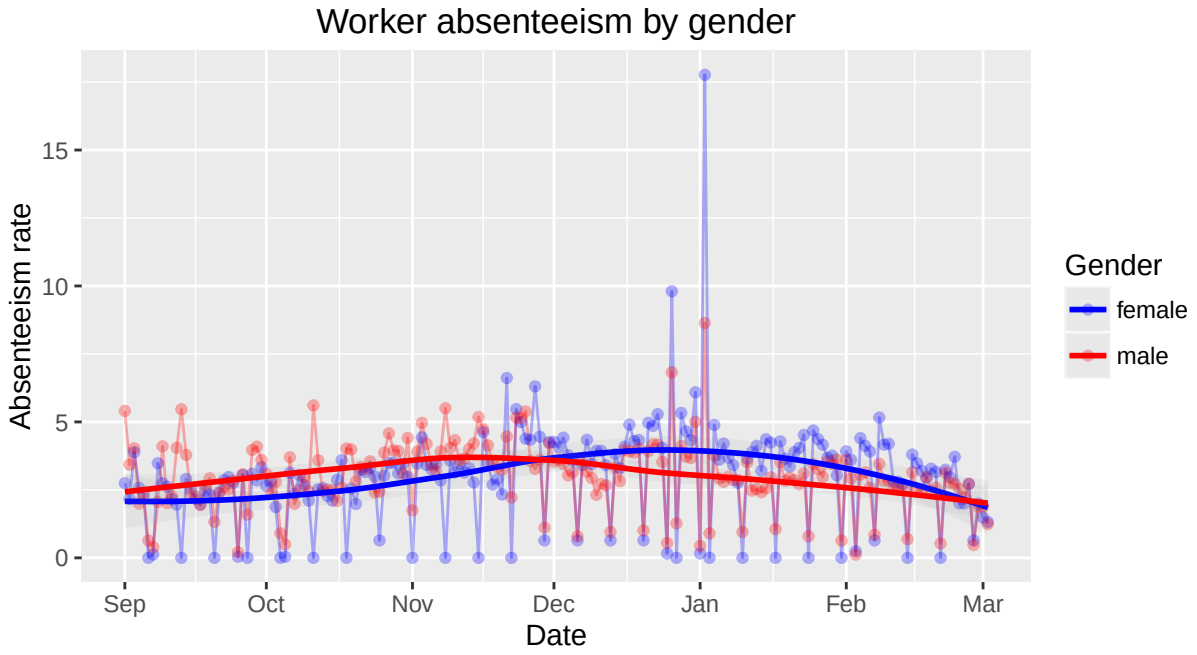


Gender

The ratio of males to females at Xinavane is greater than 2 to 1:

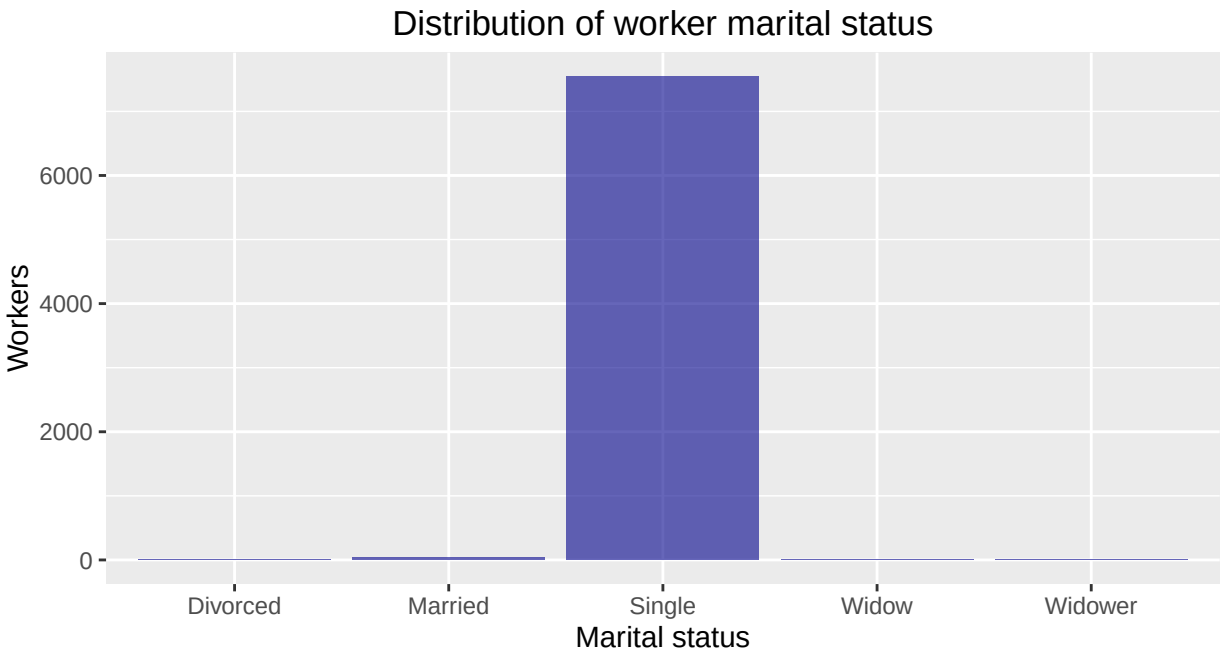


If we examine absences over time by gender, it is difficult to distinguish a trend:



Marital status

Either (a) the facility has a highly unusual subset of single Mozambicans or (b) marriage is underreported in the data:



Details

All code for the cleaning, analysis and generation of this report are hosted on github.


ISGlobal
Barcelona
Institute for
Global Health