

Xinavane data overview

Laia Cirera

Joe Brew

Elisa Sicuri

August 31, 2017

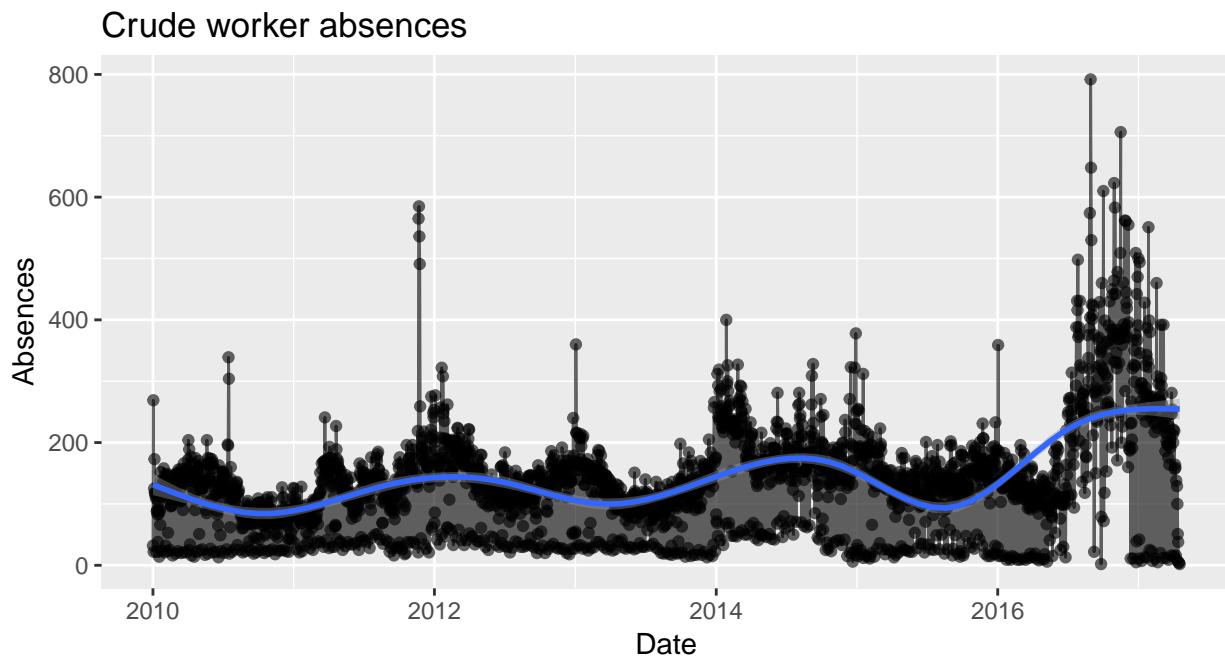
Contents

Exploration	2
Absences	2
Adjustment for worker days	2
Absenteeism rate	4
All cause absences	4
Sickness only absences	4
Absenteeism rate by worker type	5
Age	5
Gender	8
Marital status	10
Analysis	11
Difference in differences	11
Sickness only absenteeism	11
All absenteeism	13
Model	14
Details	14

Exploration

Absences

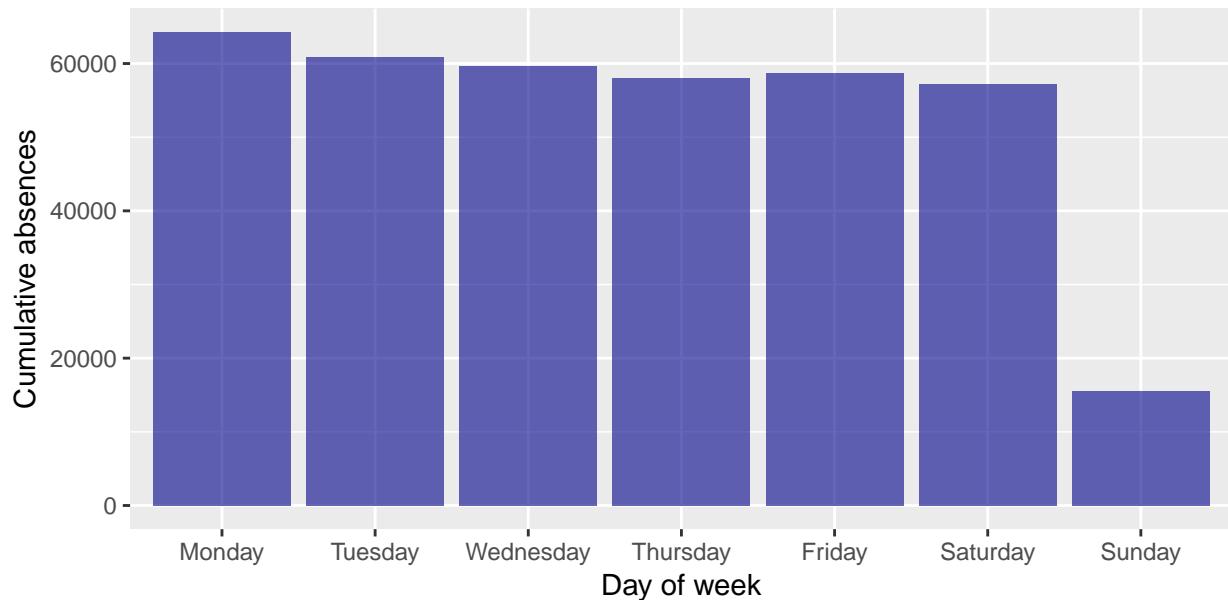
Absenteeism data spans from January 01, 2010 until April 17, 2017. On average, there are 140 absences per day. Though there appears to be some longer-term variation in the below chart, it's clear that there is annual (and even sub-annual) seasonality.



Adjustment for worker days

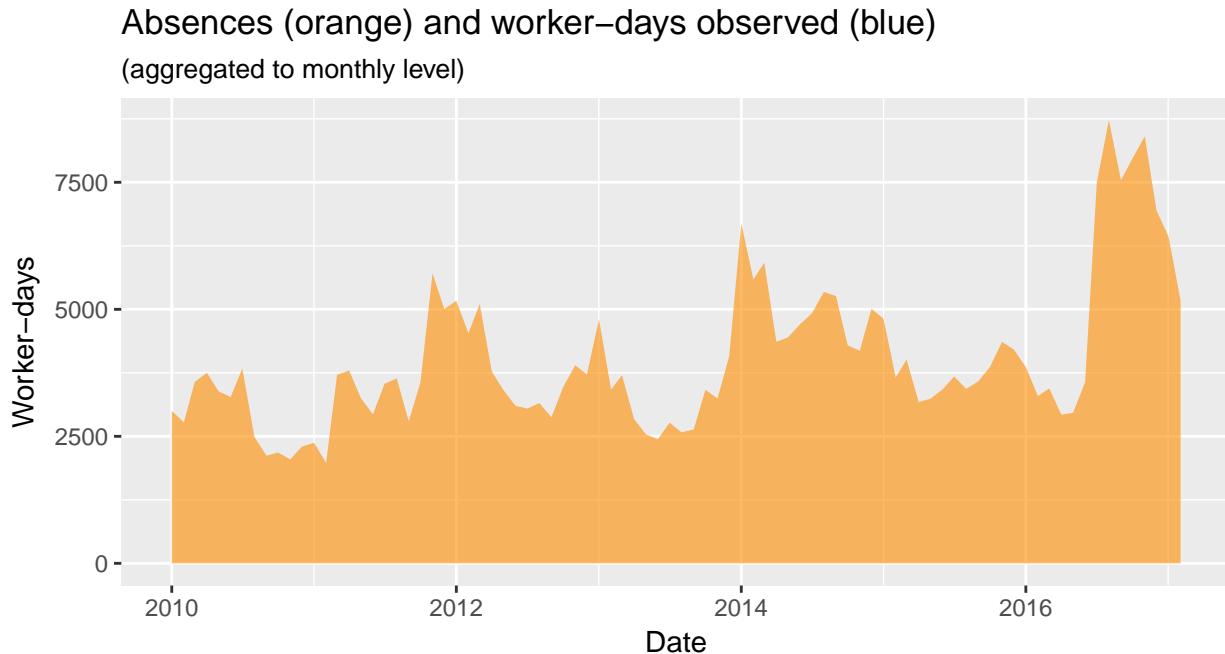
Crude absence data is relatively useless, given that it doesn't take into account the number of workers "susceptible" of absence on any given day (ie, the number of workers who were *supposed* to work). This explains why there are much fewer absences at certain times (Sunday) relative to others.

Crude absences by day of week



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, based on (a) their days designated as “off”, (b) their company entry date, and (c) their last payment date. Note, because this method does not take into account the fact that some workers go through periods of inactivity between their company entry date and last payment date (ie, seasonal workers), we have to assume “activity” during the entirety of a worker’s period of observation. This has the effect of inflating our person-time at risk (and, by extension, deflating our absenteeism rate estimates).

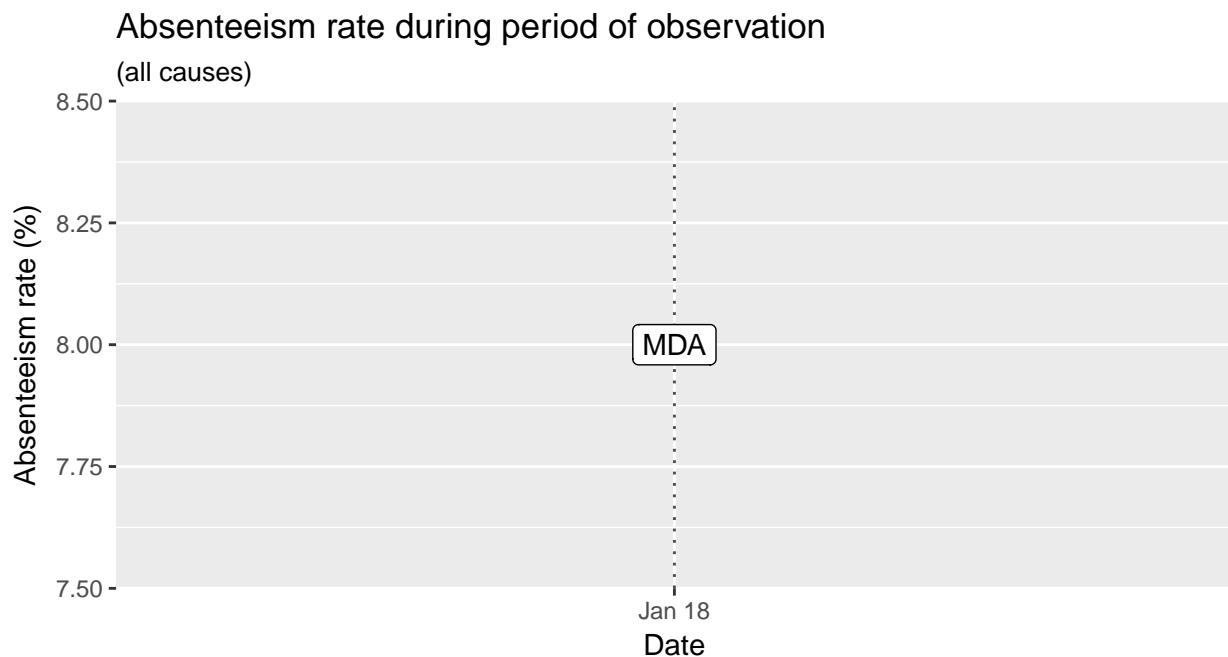
Essentially, we calculate the number of eligible workers for each day (the denominator in our rate). The result looks like this:



Absenteeism rate

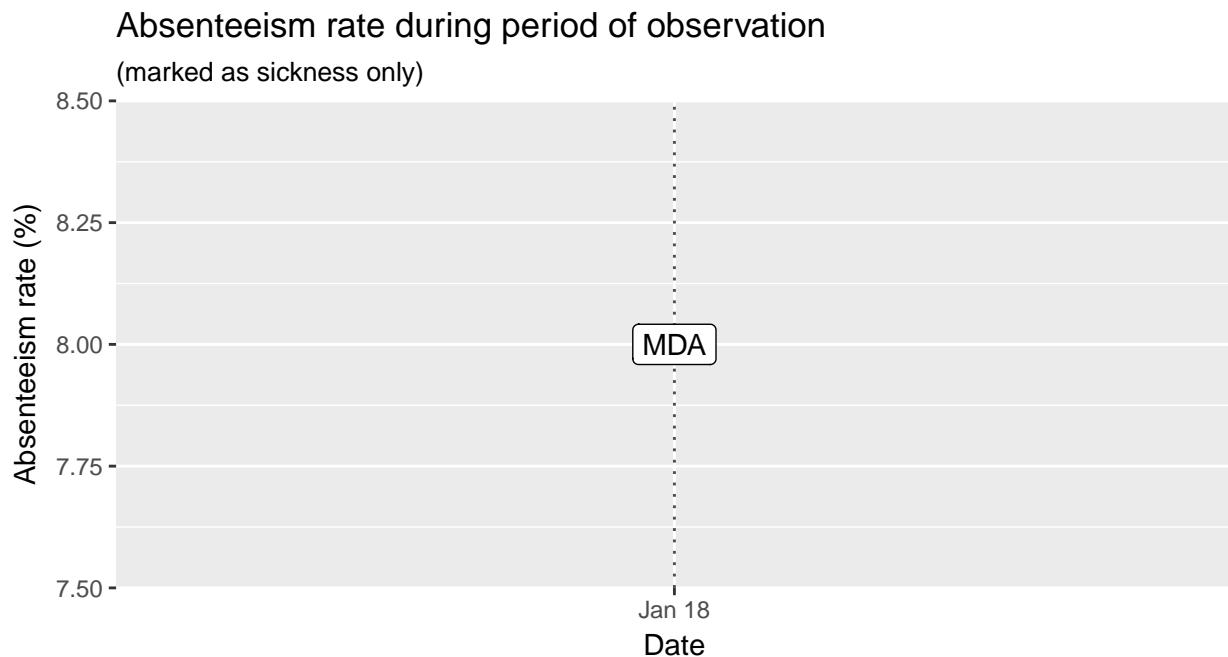
All cause absences

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 NA%.

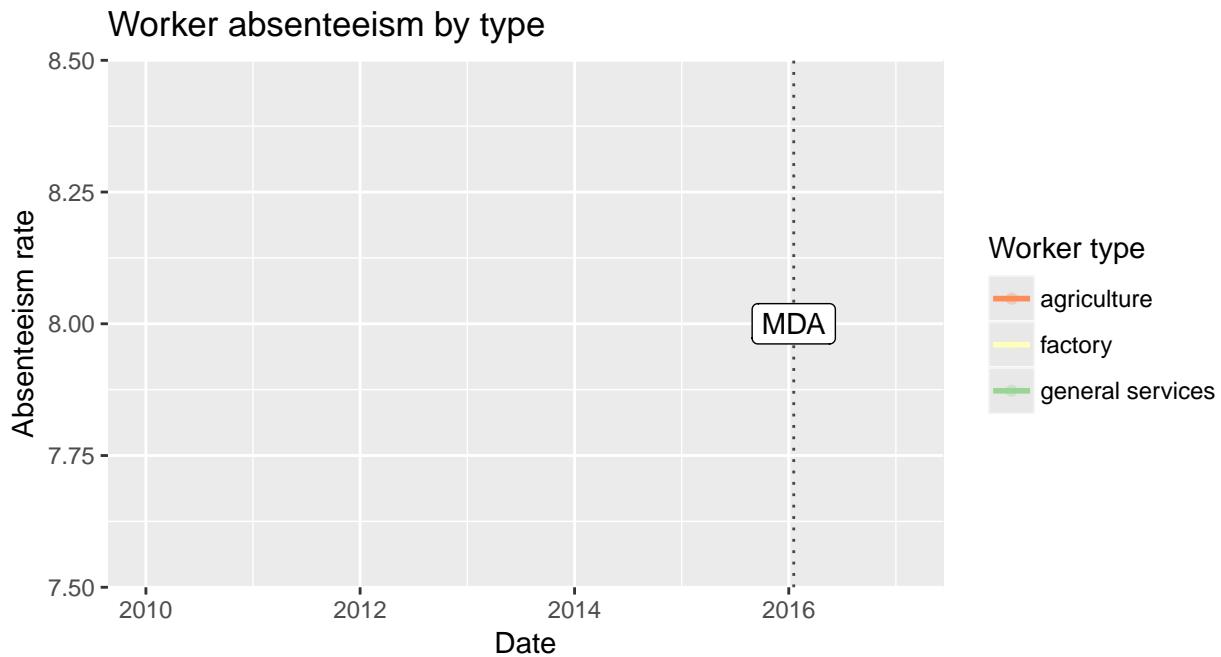
Sickness only absences



The daily average sickness-attributable absenteeism rate (adjusted for worker eligibility) is NA%.

Absenteeism rate by worker type

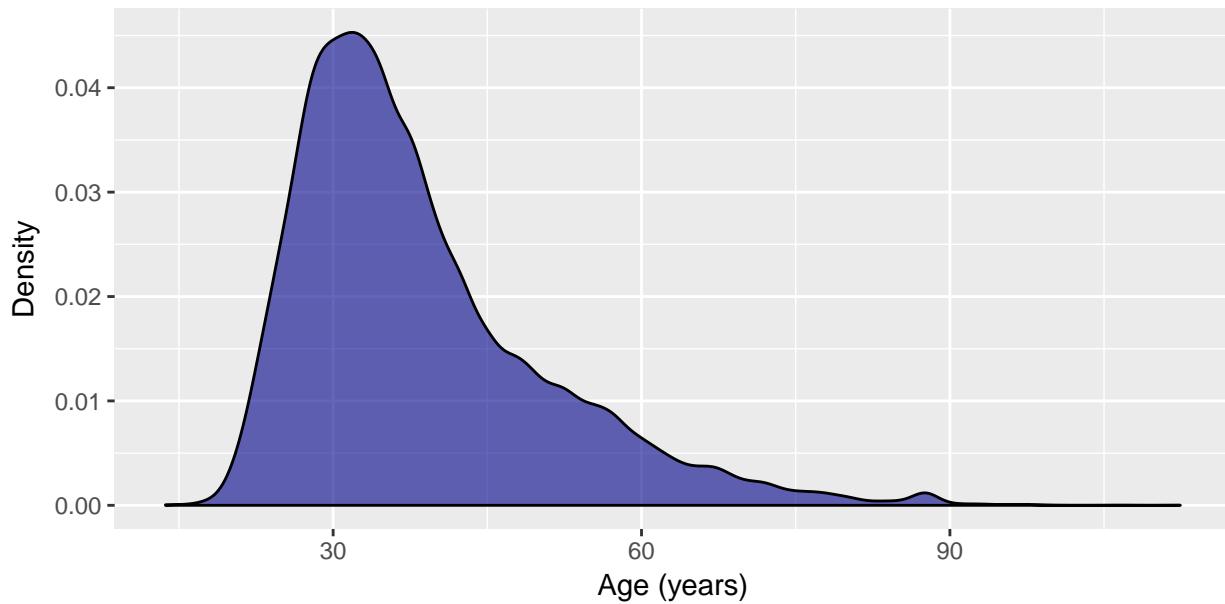
Agricultural workers have significantly higher absenteeism than those that work in the factory or in general services.



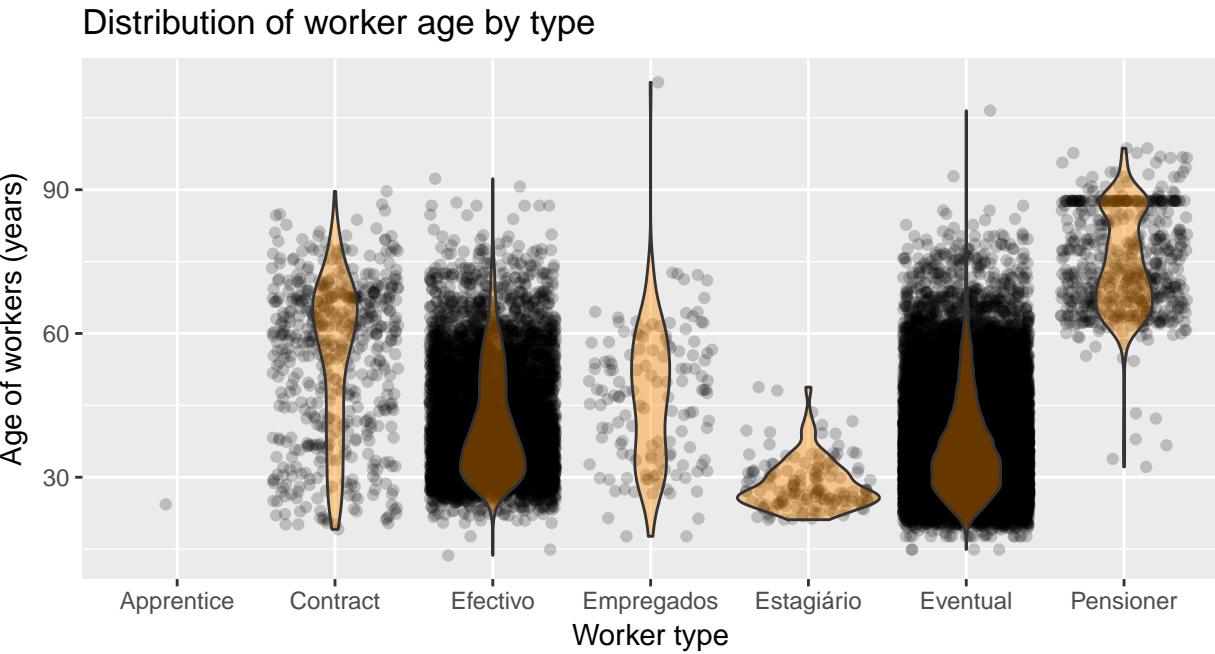
Age

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.

Age distribution of Xinavane workers



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



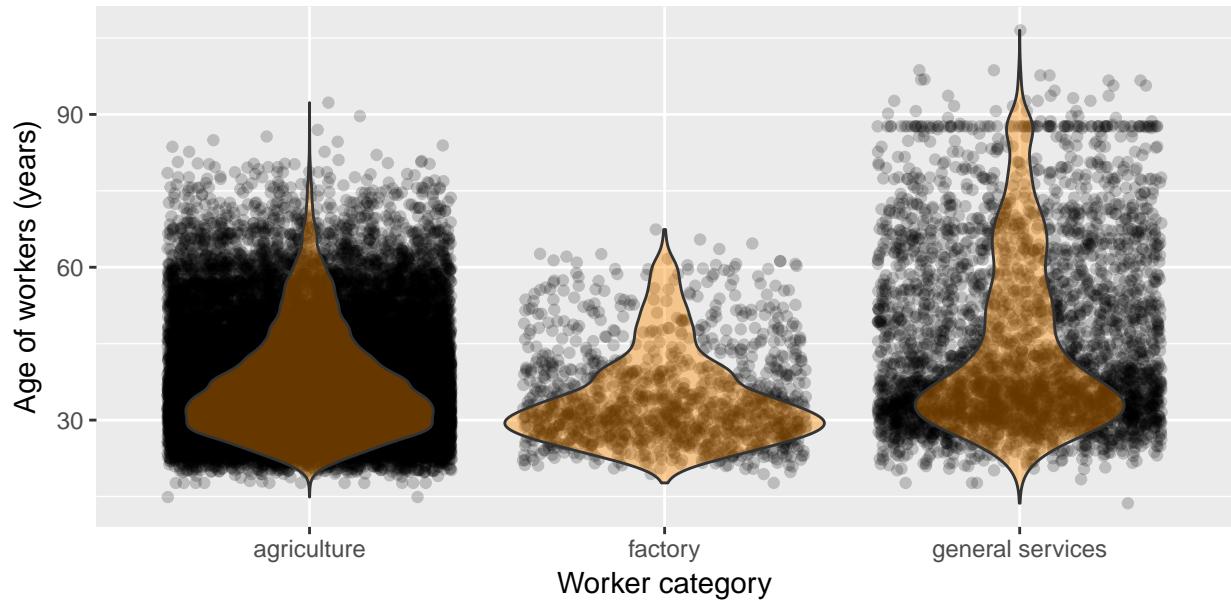
Worker age also varies significantly by category:

Distribution of worker age by category



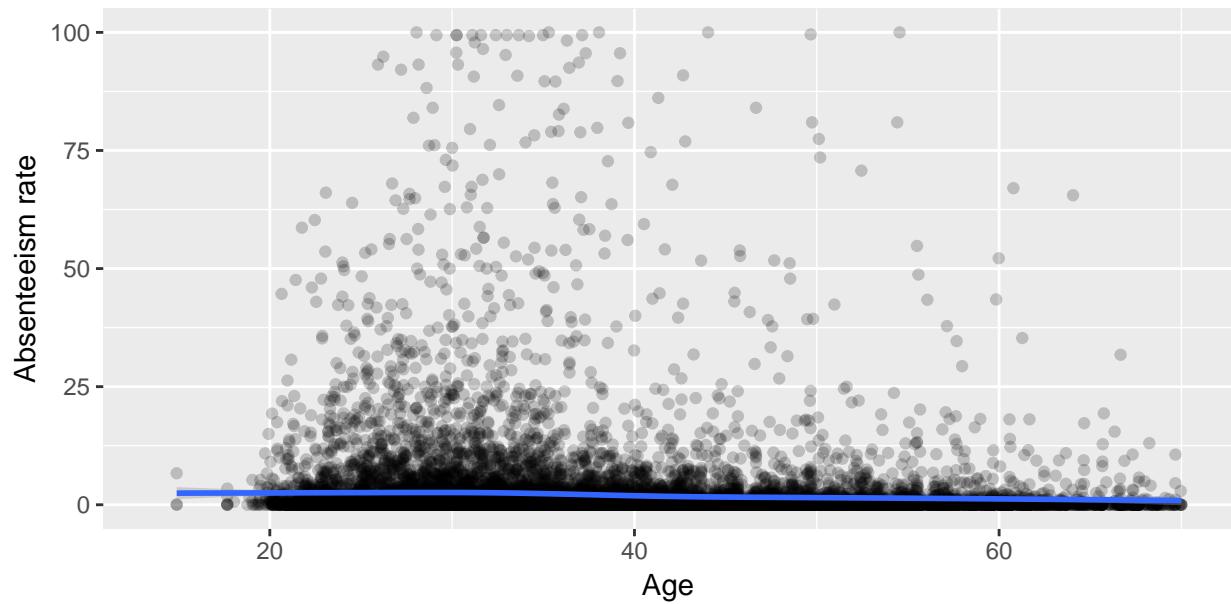
As well as by work location:

Distribution of worker age by category

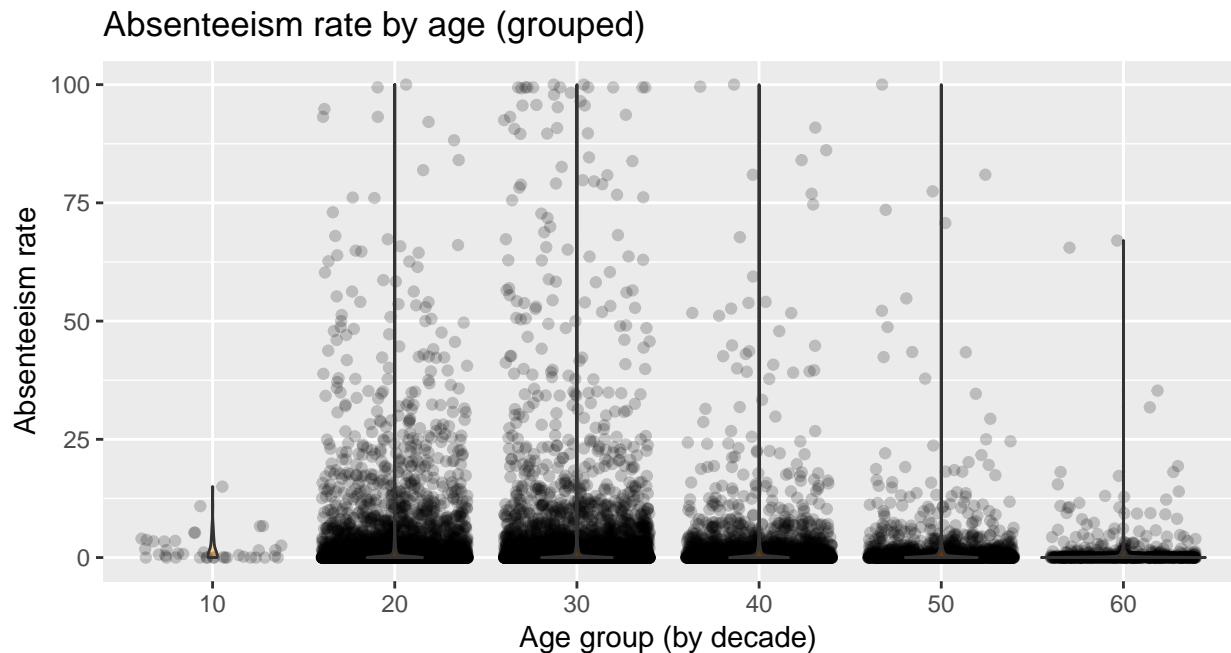


Age appears to be associated, albeit non-linearly, with absenteeism.

Age and absenteeism



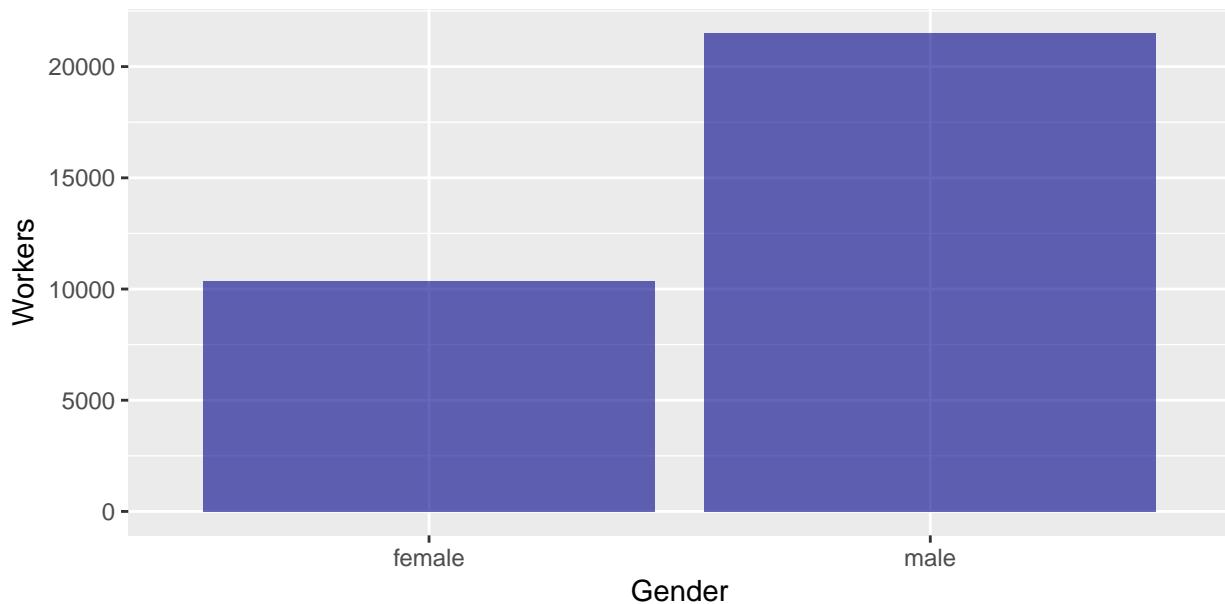
When we group by decade, a trend becomes slightly more apparent:



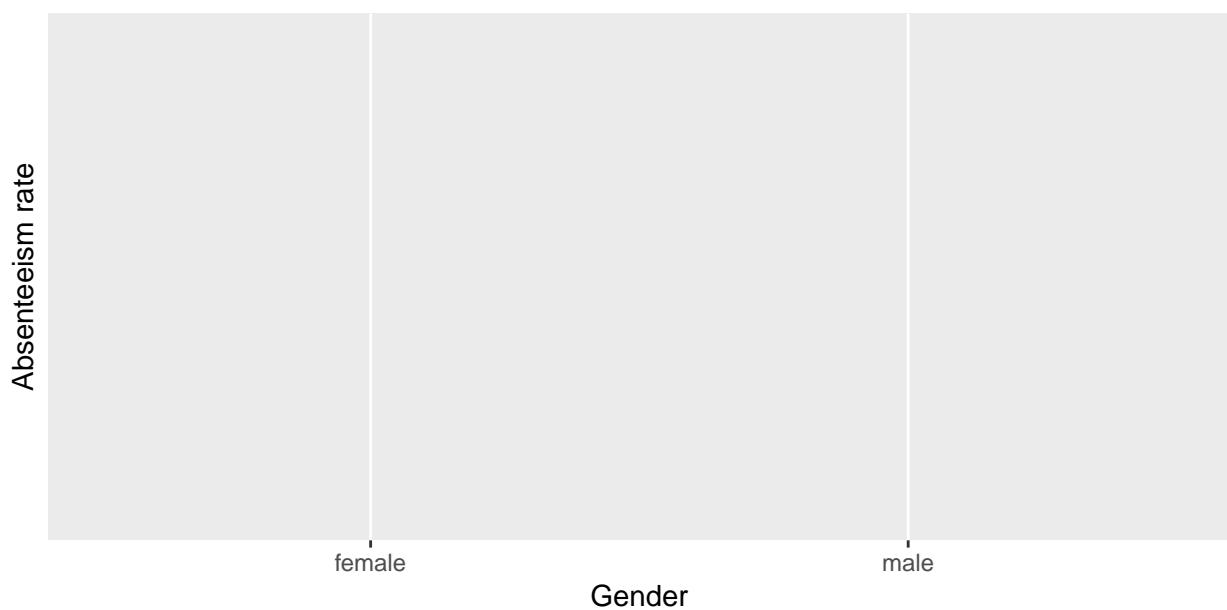
Gender

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

Distribution of worker gender

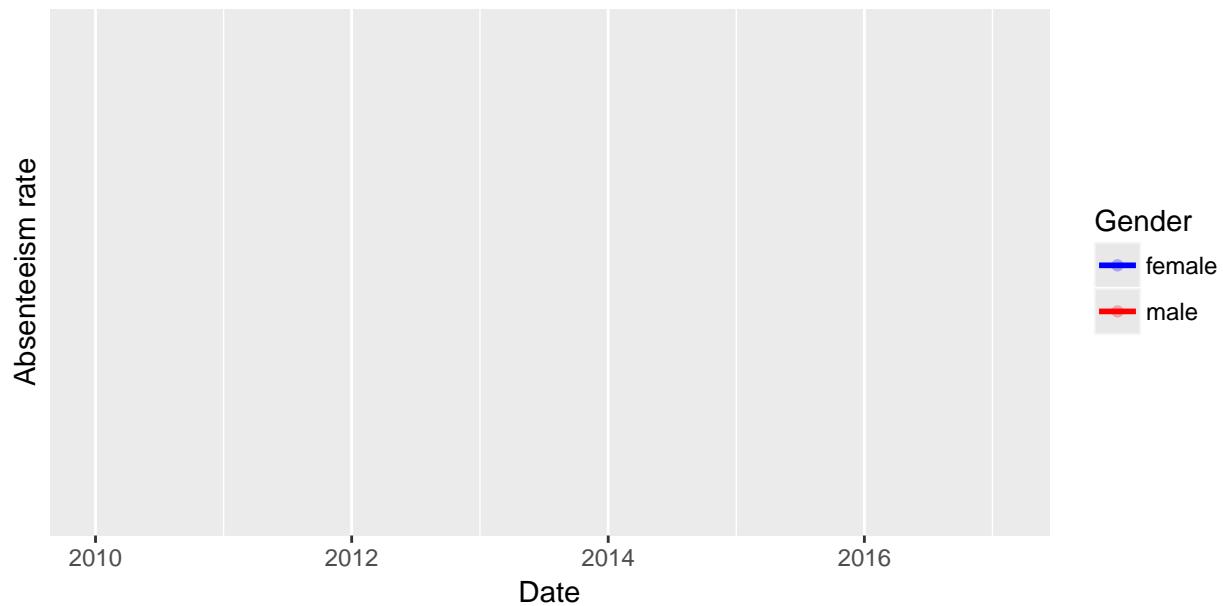


Absenteeism rate by gender



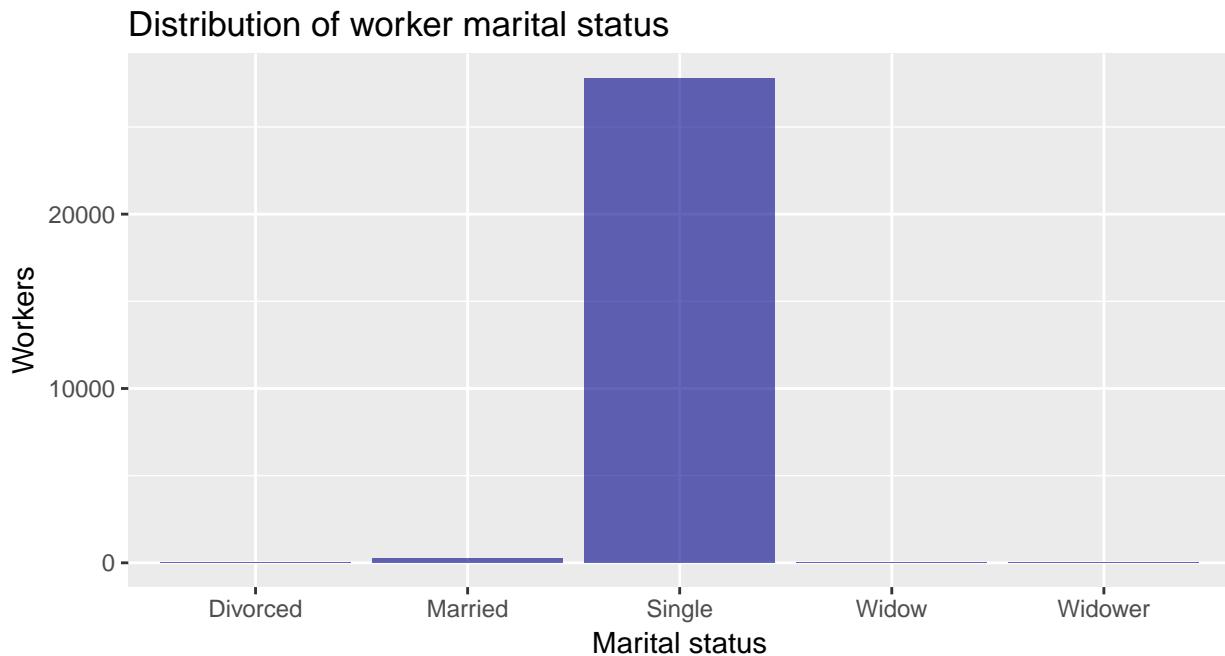
If we examine absences over time by gender, trends are apparent.

Worker absenteeism by gender



Marital status

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



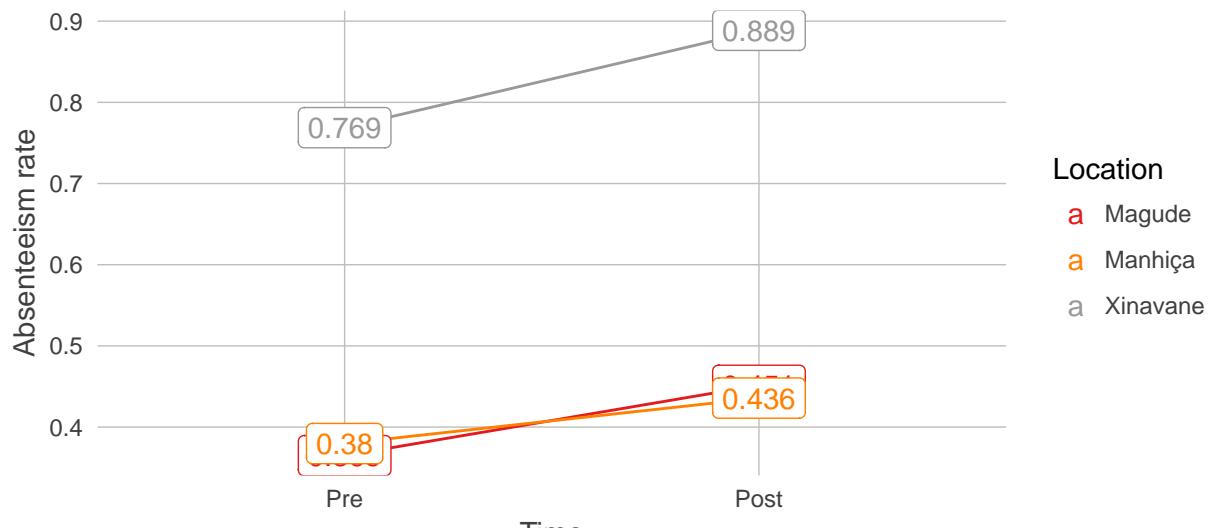
Analysis

Difference in differences

Sickness only absenteeism

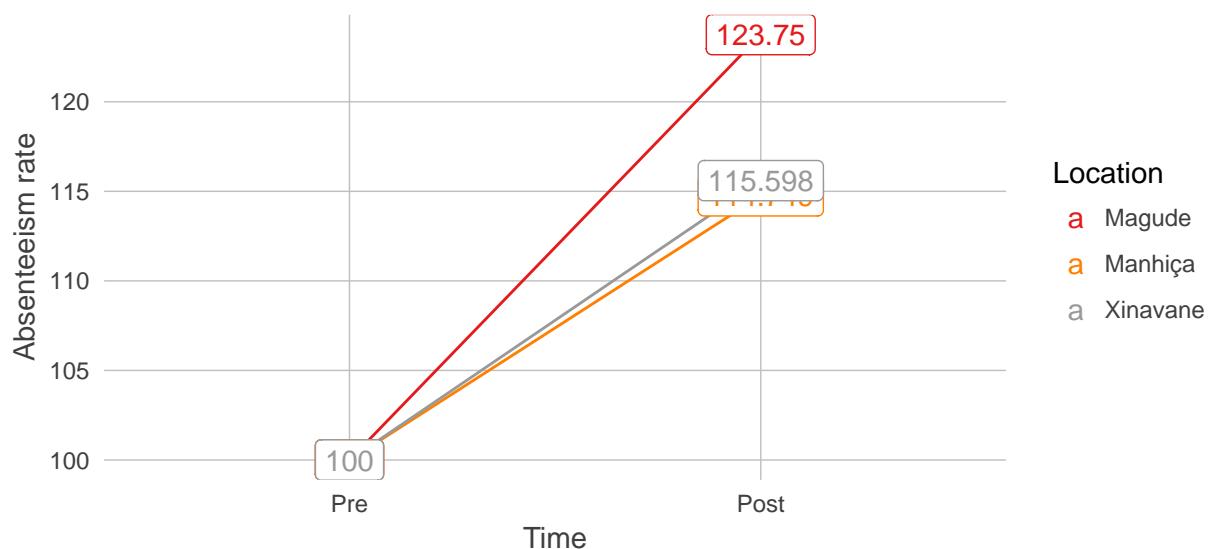
Sickness only absenteeism rate: before and after

Using company-based residential location data



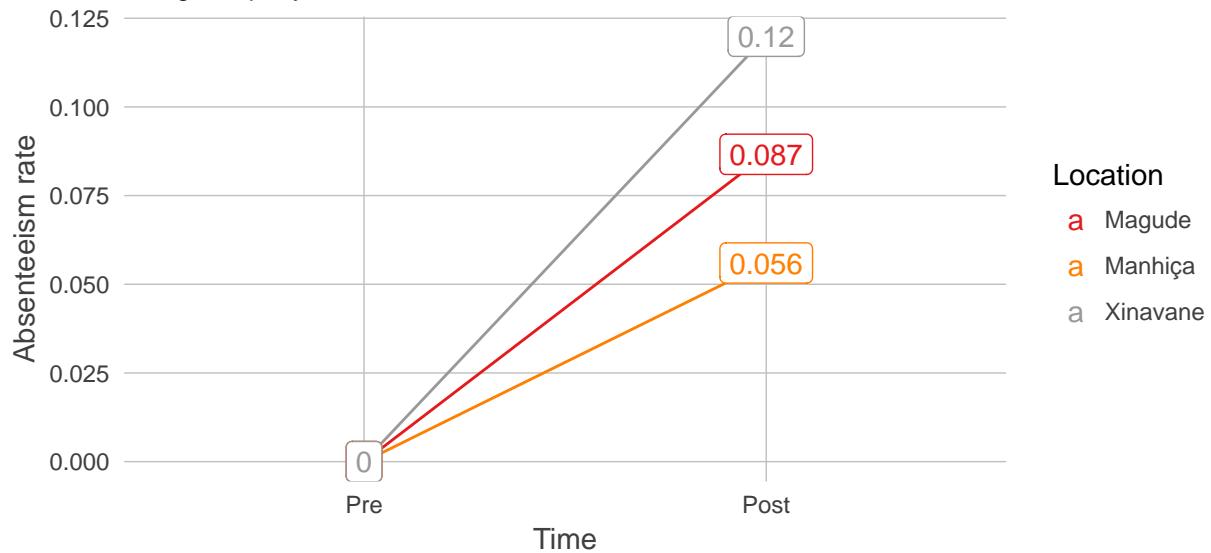
Relative scaled sickness only absenteeism rate: before and after

Using company-based residential location data



Absolute scaled sickness only absenteeism rate: before and after

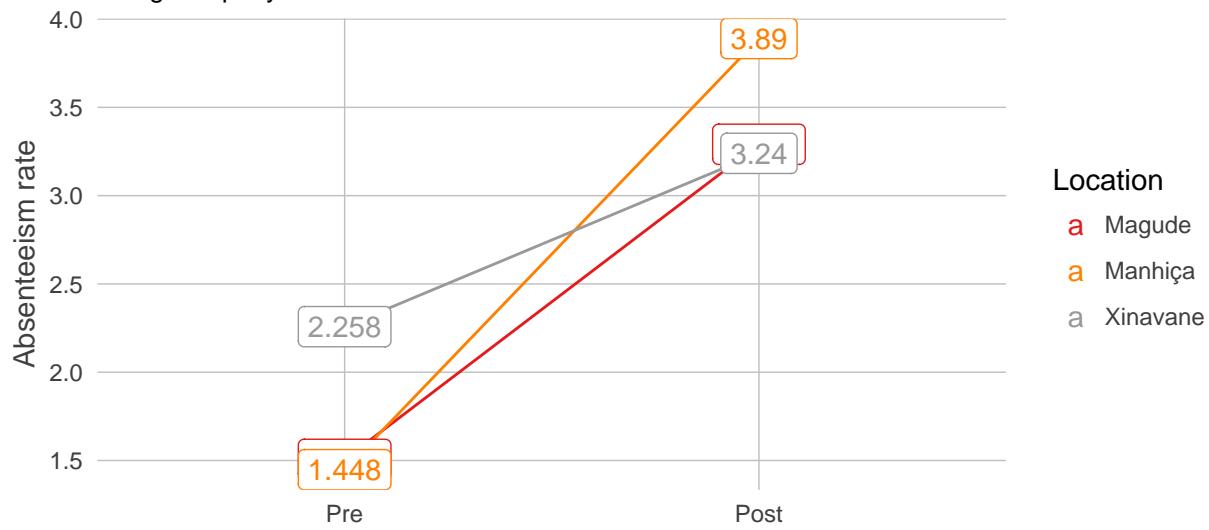
Using company-based residential location data



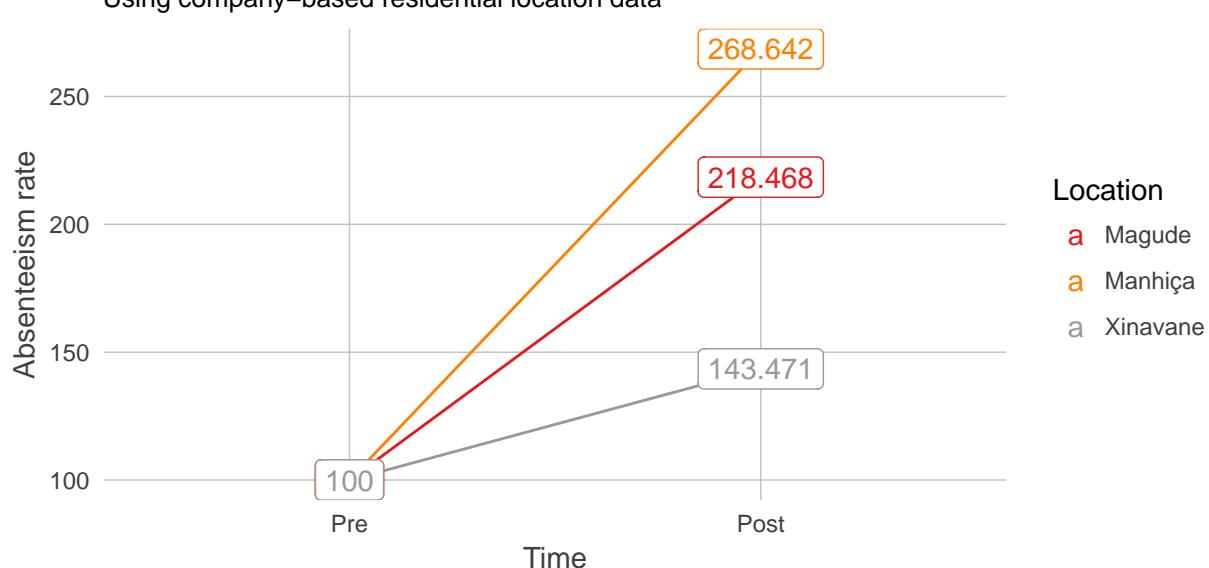
All absenteeism

All absenteeism rate: before and after

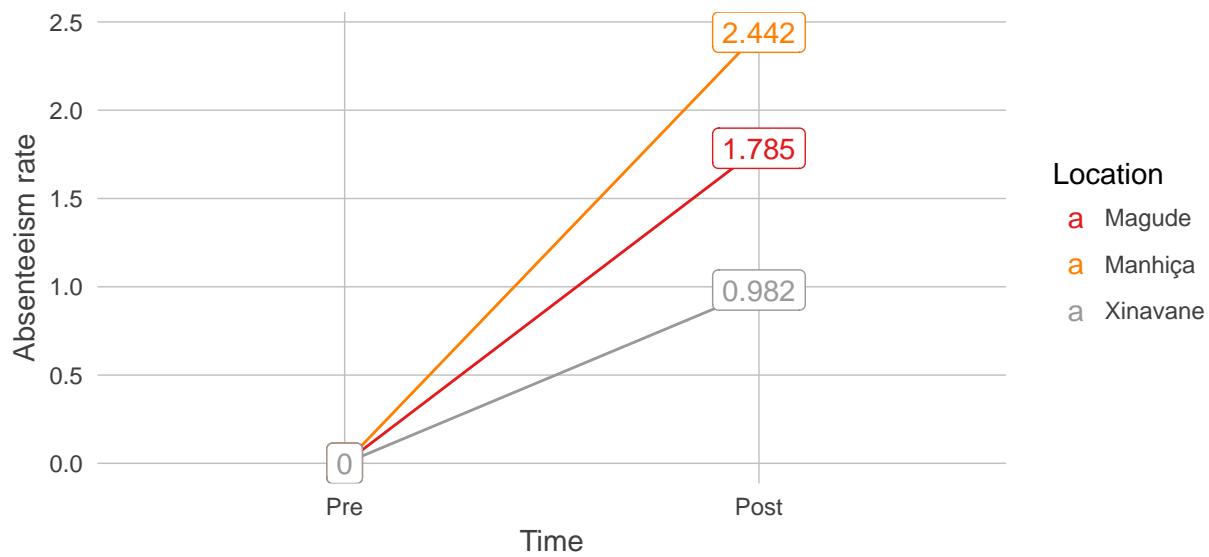
Using company-based residential location data



Relative scaled all absenteeism rate: before and after
Using company-based residential location data



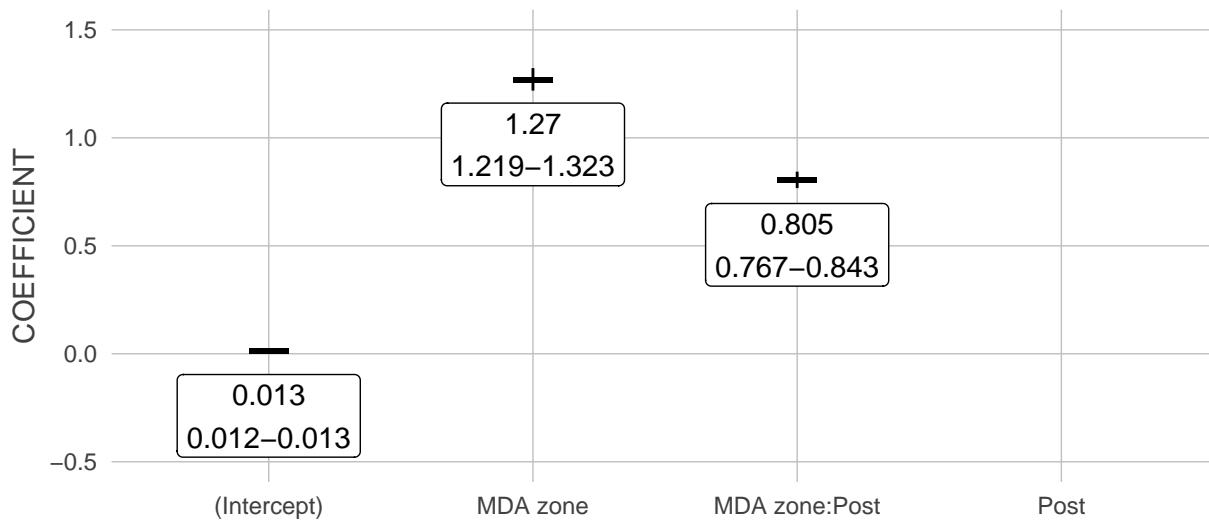
Absolute scaled all absenteeism rate: before and after Using company-based residential location data



Model

Odds ratios for absence

April 2015 to April 2016



Details

All code for the cleaning, analysis and generation of this report are hosted on Joe Brew's github page.

