# California Bay Area Zipcodes

There are gaps in the number sequence – a few spot checks showed that there are small (single residences or half city blocks) in the missing zipcodes. For the bay area cities I completed the zipcode list so every case zipcode is in the list.

# Cases with no employees or assessed penalty

30% of the San Jose Metro cases has no employees and no assessed penalties. I am working on the assumption that these are complaints that resulted in a finding of No Violation found.

# NAICS cross with SOC

Case drive on San Jose Metro Construction Industry (NAICS 23) due to my familiarity

There are NAISCS cross with SOC on BLS site (cannot find excel file) <http://www.bls.gov/oes/current/naics5_238140.htm> The NAISCxSOC example for “NAICS 238140 - Masonry Contractors” needs improvement.

Thirteen (of 42codes) NAISC codes map to multiple SOC codes – but in practice the DOL uses these thirteen codes for 55% off all cases; see next section on ‘other’ codes

* 236000 Residential and Commercial Building, (maps to 45 crafts)
* 237200 Heavy:Land Subdivision (only 3 cases, 0% of total)
* 237300 Heavy:Highway, Street, and Bridge (maps to crafts: Laborers, Carpenters, Operators, Masons, Teamsters, Ironworkers)
* Other
  + 237900 Heavy:Other (see other codes)
  + 238190 Specialty:Other Foundation, Structure, and Building Exterior (see other codes)
  + 238290 Specialty:Other Building Equipment (see other codes)
  + 238390 Specialty:Other Building Finishing (see other codes)
  + 238990 Specialty:Other Specialty Trade (see other codes)

53% of codes placed in an ‘other’ category and based on a sample audit of San Jose Metro 75% of ‘other’ codes are ‘code dumping’

* 238190 NAICS: Specialty:Other Foundation, Structure, and Building Exterior
  + San Benito Heating and Sheet Metal, Inc. 🡪miscode (238220:MEP)
  + D.H. Smith Company, Inc. (drywall) 🡪miscode (238310:drywall)
  + Vickers Concrete Sawing, Inc. 🡪 correct code
  + Jensen Landscaping 🡪 miscode(561730:landscaping)
* 238900 NAICS: Specialty:Other Specialty Trade
  + 11 cases in various sub codes
  + Looks like dumping
* Suspect that “Building: Residential & Commercial” is used like an ‘other’ category

Conclusion: 40% of construction activities are ‘dumped’ into ‘other’ and are of no clear trade

Looking at the entire state of California 55% of cases in an ‘other’ category, therefore this pattern looks to hold true generally from what I saw in the San Jose Metro

My intuition is the ‘other’ code categories are capturing workers that have no defined trade and are multicraft. They may be carpenters one day, plumbers the next, and then work as electricians. This role is tyical of non-union workers and makes categorizing the worker difficult. Some specialty trades will have a defined type of work they typically bid and these most likely account for the other 45% of workers.

Bad guys are just bad guys at this point: 0s all across the board will be compliance cases, no violations and Could be any Act though, family leave, H-1B, etc. I was just looking at a ballpark if there was four or twelve - I started with construction because I know the industry well enough that I can tell if the charts don't make sense. Once I get the process down then I can replicate for all the other industries and assume it worked for construction then the results should be valid for the others. I think the main takeaway is there is a huge 'other' category to be looking for when we move over to care homes or restaurants. Zeroing on the workers that transit between occupations as a part of their occupation is a class of worker in itself.

To validate this intuition I suggest interviewing a sample of violators and ask about the workers craft category.

Figure 1 The ratio of violations by work type generalizes and scales from the San Jose metropolitan region to the entire state

Using construction as a case study: I could map 40% of the DOL cases from NAICS to SOC format, the other 60% map from 1 NAICS code to many undefinable SOC codes. For example, NAICS:Residential\_Construction maps to 45 SOC codes such as electrician, roofer, framer, mason, etc. 30% of the DOL dataset is essentially coded as 'Residential & Commercial Construction. In the best case both codes should be used by the DOL so they have a code looking like 23.6117-47.4031 for Building:New Housing Fence Erectors. You could find these workers by visiting a new housing subdivision development and look for the guys building the fences. There is a good chance these guys are paid in cash, are not paid overtime, and are not having contributions made to their benefits. Knowing just 23.6117 or 47.4031is not enough to pinpoint fence erectors on subdivisions. Knowing just 23.6117 only tells me new housing and this can mean anything from fence erectors to roofers and electricians. Knowing fence erectors is more helpful since they cross multiple domains of NAICS codes.

One thing I found - in construction, half the cases are basically NAICS coded to 'other.'  For example they have residential construction but this category of worker does not exist. The electrician on a residential project (NAICS 236) is just as likely to drive next to install some specialty equipment (NAICS 238) maybe on the same project where they would be doing NAICS 236 work. It makes it a hack assigning SOC employee population.

I suspect the residential/commercial building and the various 'other' categories are a catchall for 'multicraft' workers; they are electricians today, plumbers tomorrow, and do the roof at the end of the project. This is fairly common and difficult to then pinpoint the craft; I created a new category for 'multi-craft.'.

# Mismatch in comparison between current BLS population and DOL violations over 20 years.

[average DOL data over 20 years for comparison ratios with current population. For example, the violations affecting pointers, take the mean number of painters affected by cases each year and compare this to the current population of painters ( 100 painters over 20 years is 5 painters per year; 5/10,000 painters is the ratio we use].

Each row in the main whd\_whisard dataset on the DOL data enforcement website is defined as "all concluded WHD compliance actions since FY 2007”. I have the following questions:

* I (Ash) see that the gap between “findings start” and “findings end” date range from 1985 to 2015 (there are some errors in date entry). My concerns are with matching up the numbers by NAICS code across years. An industry could have violations that went as far back as the 80s, but if we compare, for example, “EEs employed in violation” with external datasets from the SUSB or CBP (e.g. “total employees in industry group”) then isn’t it somewhat misleading? Say we’re using 2014 industry aggregate numbers by NAICS code. Then something like “Backwages per employee” when compared to aggregate 2015 numbers doesn’t seem right because these back wages were owed over a 25-30 year span. We’re essentially comparing past violations to current numbers to create any ratios, and this could be misleading in subsequent analyses. Does this make sense? I’d love to hear your thoughts.
* Assuming this is a cause for concern, a possible solution could be to take the average of industry statistics over the same 25 or so year span, and then calculate any ratios against industry numbers. This sort of analysis, in essence, will discount any variations in an industry group’s violation patterns over time.
* FP: For California the first cases are 2002 and the last are end of 2015; there are a few earlier cases but these look like investgations that started in 2002 or 2003 with an earlier violation start date; assumed based on number of cases in 2002 (111) versus 2001 (37), each preceding year has vastly fewer cases. This is 15 years.
* Averaging the violations by 15 years seems too much. In the San Francisco-Redwood City-South San Francisco, CA metro region there are 425 construction (div 23) cases. Currently there are 30,000 construction workers in this region. Just due to retirements we can assume half those 30,000 workers were in construction 15 years ago, the other half retired. Construction has a higher velocity of turnover entering and exiting the industry (citation, xxxx) due to the rigorous conditions. There for we can assume 15 years is excessive. My intuition is six to twelve years is the mean construction career (citation, xxxx). Of the 30,000 current workers, I would assume three years ago the majority were the same 30,000. Six years ago, man are but before six years ago, thirty percent are probably new workers.
* I propose we take the mean citations as over six years. For example, 425/6 = 70 and then 70/30,000 = 0.2% of workers in this metropolitan area have been part of a DOL case. 425/15 = 28 violations and so 28/30,000 0.1% violations per worker seems too much of a discount. I agree that 425/30,000 = 1% of workers have been part of a DOL case is unrealistic given this is over 15 years. A six year interval seems fair.
* Ash thought comparing 15 years of cases to current population is a mismatch; I think he is using the mean cases per year, I am using mean cases for 6 years. So, mean cases per 6 years normalized by # of workers in that region. We are ignoring population growth. The idea is the construction worker standing here today was not the same dude standing there 15 years ago so a DOL case from 10 years ago means squat to his working conditions
* My six year intuition is based on: From my LiUNA survey, the average Union laborer has 10 years. 1/3 are under five years and 2/3 are over 5 years. The Laborers are thought to have the highest velocity in turn over simply because it is difficult. Non-Union usually have a higher velocity in turnover, if the Laborers are the benchmark then non-union are probably on average less than 10 years in a trade before they move on. The foreman and supervisors have more longevity (the Judas fucks they are)  but they know the rest will turn over every ten years. I think I have seen supervisors that assume they can get away with it because they know you won't be around in another five years. If there was a recent action, a few years, they shape up and treat us better.

# Checks

* DOL case where these is no BLS population – none found in Bay Area Metro regions

# Construction Bay Area Dive: Heatmap by Region and Occupation

Mean\_wage = average of mean wage and median wage

DOL\_ratio = ratio of DOL\_annual\_case to regional occupation population

DOL\_annual\_case = DOL cases total divided by 6 (years)

Heat Index = b ((100-mean\_wage)/100) + b (regional\_ee/10,000) + b (1-(DOL\_ratio) )

To clean up the duplicate NAICS to SOC codes I went through and by hand looked for NAICS codes violations that have no SOC coded population – indicating that the NAICS is duplicate (can’t have a violation without workers).

Missed data

* Salinas BLS dataset is missing but has DOL cases
* Napa has BLS but no DOL cases