Visualising growth in occupations in one industry

A chart is doing the rounds purporting to show the number of administrators working in health care in the USA has grown much faster than the number of physicians – more than 3,000% growth from 1970 to 2009 for administrators (allegedly) compared to about 90% or so for physicians. I don't much like the original chart so I've relegated it to the bottom of this post. It presumably dates from the time of the debates about the introduction of the Affordable Care Act (aka 'Obamacare'). I find it very difficult to believe the 3,000% number, and suspect there is either deliberate definitional sleight of hand going on, or a genuine classification challenge. One obvious possibility is that some "administrator" classification has been cherry-picked that was very rarely present under that name in the 1970s, and much of the growth is movement from other differently-classified roles into that one.

A similar visualisation with Australian Labour Force Survey data

It did cross my mind that the problem was the visualisation method; in fact the tweet that brought this to my attention was from a researcher wondering what it would look like if it showed absolute numbers rather than cumulative growth. While I'm not really interested in the facts of "administrators" in the US health system, the broader data viz question sounded like something I should know about. So I had a look at Australian figures from the Australian Bureau of Statistics' Labour Force Survey Quarterly Detail. Here is my own version of a chart showing cumulative growth in various occupations in an industry:

Actually, I think my chart is much better than the US original, not only because it uses an official and well-defined occupation classification, but because it has a go at showing absolute size as well. So we can see that while the total hours worked in the health care and human services industry by managers and professionals who aren't health-specific (more on this below) have grown fast, the orange and grey dots are still small compared to the pink dots that represents health professionals.

The industry I'm looking at here is "Health Care and Social Assistance", so some of those managers and other professionals (lawyers, accountants, statisticians, etc) are in social assistance rather than health, but this is as granular as we can get for an occupation and industry crosstabulation with this data without a custom request to access the microdata.

In fact, clearly one of the big stories from this chart is the thick blue dots and the rapidly rising blue line – community and personal care workers. The biggest occupations by far in that category in this industry are "Aged and Disabled Carers" and "Child Carers", and the growth in importance in these roles (particularly the former) is one of the big news stories of the economy seen over a few decades' perspective.

I have split the "Professionals" ANZSCO code (the lowest level published by industry in this dataset) into health and other, by getting employment hours for everyone in occupation codes 2500 to 2544 ("Health Professionals not further defined" to "Registered Nurses"). This is from a different cut of the data from the industry version and is only published for "all industries". I adjust the all-industries health professionals number downwards by about 14%, based on the 2011 Census which tells us that there were 433,726 health professionals in total of which 373,609 worked in the health care and social assistance industry (see screenshot at bottom of post). For example, a mining company or sports team can hire a doctor or nurse. To avoid working on this thing all weekend, I've applied that single correction across all years of data.

Other ways of showing this data

What are the other obvious ways to visualise this data? Obviously, in absolute numbers as a stacked area chart:

… or as above but with position "fill" so we see changing proportions:

All three of these methods are completely valid.

- The very first line chart – cumulative growth – is visually equivalent to converting labour hours to an index. It's great for showing growth over time, and for many purposes would be suitable. For example, it nicely highlights that the number of labourers in the health care and social assistance industry has

declined, and the fastest growing occupation types are managers and non-health professionals.
- The second – absolute numbers – highlights the aggregate size and growth of labour in these occupations, while still allowing basic comparisons of changes.
- The third – proportions – lets you see changes in the proportion of the workforce while still getting a snapshot overview (like a pie chart would, but for many times). In this case the change we see is the growth in community and personal service workers rather than health professionals.

Physicians / medical practitioners compared to other health professionals?

The original US chart had focused specifically on "physicians" and I've used a broader category of "Health Professionals". This prompted me to do one last bit of analysis with this – to find out how much of the Australian health profession's labour is by medical practitioners and whether this is changing. I was surprised to see that the proportion of all health professional labour done by medical practitioners of various sorts (there is no "physician" in the ANZSCO so I chose the combination of unit groups I thought was closest to this) has stayed pretty constant over the past 35 years:

That's a very boring chart, but it's boring for interesting reasons – a fairly steady composition of employment hours within the health professionals category, at least when divided between medical practitioners and others.

So what occupations are growing fast?

Finally, I was intrigued by the 3,000% cumulative growth in the one occupation in the original US chart. Could an occupation really grow this much in a few decades? Turns out we have a couple in the Australian data, but in our case I think these are genuine changes in employment patterns. Professional outdoor guides and ICT test engineers are two professions that we believe really have grown materially in the last few decades, partly due to changes in demand and workflow and partly due to specialisation and reclassification of other roles.

I like that plot because it gives a real snapshot (at least from one perspective) of how the economy has changed over 32 years.

Code

Here's today's R code, all in one chunk. The most interesting thing here is the need to use the occupation detailed data (cube EQ08) to separate out the single digit occupation data that we get from the higher level industry by occupation data in cube EQ09.

```
library(tidyverse)
library(readxl)
library(janitor)
library(scales)
library(RColorBrewer)
library(lubridate)

#==================Data management=======================

#----------Industry by occupation----------------
url <- "https://www.abs.gov.au/AUSSTATS/subscriber.nsf/log?
openagent&eq09.zip&6291.0.55.003&Data%20Cubes&0A68E700485EF985CA2585910016AF
26&0&May%202020&25.06.2020&Latest"
download.file(url, "lfs-eq09.zip", mode = "wb")
unzip("lfs-eq09.zip")

eq09 <- read_excel("EQ09.xlsx", sheet = "Data 1", skip = 3) %>%
  clean_names() %>%
  rename(industry1 = industry_division_of_main_job_
anzsic_2006_rev_2_0,
         occupation1 = occupation_major_group_of_
```

```
main_job_anzsco_2013_v1_2) %>%
  mutate(total_hours =
           number_of_hours_actually_worked_in_all_jobs_employed_
full_time_000_hours +
           number_of_hours_actually_worked_in_all_jobs_employed_
part_time_000_hours)

#------------Detailed occupation-------------
url <- "https://www.abs.gov.au/AUSSTATS/subscriber.nsf/log?
openagent&eq08.zip&6291.0.55.003&Data%20Cubes&
CB6124B8CB5B515DCA2585910016A499&0&May%202020&25.06.2020&Latest"
download.file(url, "lfs-eq08.zip", mode = "wb")
unzip("lfs-eq08.zip")

eq08 <- read_excel("EQ08.xlsx", sheet = "Data 1", skip = 3) %>%
  clean_names() %>%
  rename(
    occupation4 = occupation_of_main_job_anzsco_2013_v1_2,
    total_hours = number_of_hours_actually_
worked_in_all_jobs_000_hours
  )

health_profs <- c(
 "2500 Health Professionals nfd"
,
 "2510 Health Diagnostic and Promotion Professionals nfd"
,
 "2511 Nutrition Professionals"
,
 "2512 Medical Imaging Professionals"
,
 "2513 Occupational and Environmental Health Professionals"
,
 "2514 Optometrists and Orthoptists"
,
 "2515 Pharmacists"
,
 "2519 Other Health Diagnostic and Promotion Professionals"
,
 "2520 Health Therapy Professionals nfd"
,
 "2521 Chiropractors and Osteopaths"
,
 "2522 Complementary Health Therapists"
,
 "2523 Dental Practitioners"
,
 "2524 Occupational Therapists"
,
 "2525 Physiotherapists"
,
 "2526 Podiatrists"
,
 "2527 Audiologists and Speech Pathologists \\ Therapists"
,
 "2530 Medical Practitioners nfd"
,
```

```
 "2531 General Practitioners and Resident Medical Officers"
,
 "2532 Anaesthetists"
,
 "2533 Specialist Physicians"
,
 "2534 Psychiatrists"
,
 "2535 Surgeons"
,
 "2539 Other Medical Practitioners"
,
 "2540 Midwifery and Nursing Professionals nfd"
,
 "2541 Midwives"
,
 "2542 Nurse Educators and Researchers"
,
 "2543 Nurse Managers"
,
 "2544 Registered Nurses"
)

#----------------------mucking around with classifications-----------
# We will find the total hours by health professionals (codes
above)...
health_profs_hours <- eq08 %>%
  filter(occupation4 %in% health_profs) %>%
  group_by(mid_quarter_month) %>%
  summarise(health_prof_hours = sum(total_hours),
            # adjust downwards to crudely remove health professionals
in other industries
            # Source: Table Builder for Census 2011 (note this ratio
is applying to our
            # whole time period, so this is really rough)
            health_prof_hours = health_prof_hours * 373609 / 433726
            ) %>%
  mutate(occupation1 = "Professionals")

# And subtract it from the Professionals in the Health Care and Social
Assistance Industry,
# to get a data frame that has two types of professionals. Note this
is problematic because
# of health professionals in other industries.
separated_profs <- eq09 %>%
  filter(industry1 == "Health Care and Social Assistance") %>%
  group_by(occupation1, mid_quarter_month) %>%
  summarise(total_hours = sum(total_hours)) %>%
  inner_join(health_profs_hours, by = c("mid_quarter_month",
"occupation1")) %>%
  mutate(other_profs_hours = total_hours - health_prof_hours) %>%
  ungroup() %>%
  select(-total_hours, -occupation1) %>%
  gather(occupation, total_hours, -mid_quarter_month) %>%
  mutate(occupation = case_when(
    occupation == "health_prof_hours" ~ "Health Professionals",
    occupation == "other_profs_hours" ~ "Non-Health Professionals"
```

```
  ))

# join back to the original data
d <- eq09 %>%
  filter(industry1 == "Health Care and Social Assistance" &
occupation1 != "Professionals") %>%
  rename(occupation = occupation1) %>%
  group_by(occupation, mid_quarter_month) %>%
  summarise(total_hours = sum(total_hours)) %>%
  ungroup() %>%
  rbind(separated_profs) %>%
  mutate(occupation = fct_reorder(occupation, total_hours))

#==========Plotting============

#----------named palette and caption-------------

occ_palette <- brewer.pal(length(unique(d$occupation)), "Set1")
names(occ_palette) <- unique(d$occupation)

the_caption <- "Source: ABS Labour Force Survey EQ08 and EQ09,
analysis by http://freerangestats.info"

#------------Line chart------------
# This is equivalent to an index - showing cumulative growth
d  %>%
  mutate(yr = year(mid_quarter_month)) %>%
  group_by(yr, occupation) %>%
  summarise(total_hours = mean(total_hours)) %>%
  group_by(occupation) %>%
  arrange(yr) %>%
  mutate(growth = total_hours / total_hours[1] - 1) %>%
  ungroup() %>%
  mutate(occupation = fct_reorder(occupation, -growth, .fun = last))
%>%
  ggplot(aes(x = yr, y = growth, colour = occupation)) +
  scale_y_continuous(label = percent) +
  geom_point(aes(size = total_hours / 1000)) +
  geom_line(stat="smooth", method = "loess", span = 1/2, alpha = 0.5,
size = 2) +
  scale_colour_manual(values = occ_palette) +
  scale_size_area(label = comma_format(accuracy = 1)) +
  theme(legend.position = "right") +
  labs(x = "",
       y = "Cumulative growth in hours since 1986",
       size = "Million hours per quarter",
       colour = "Occupation",
       subtitle = "Hours by occupation of workers in Australia's
health care and social assistance industry",
       title = "Strong growth in managers and non-health
professionals, but absolute numbers are small",
       caption = the_caption)



#----------stacked and filled area charts----------------
# fundamental guts of the plot with no geom
```

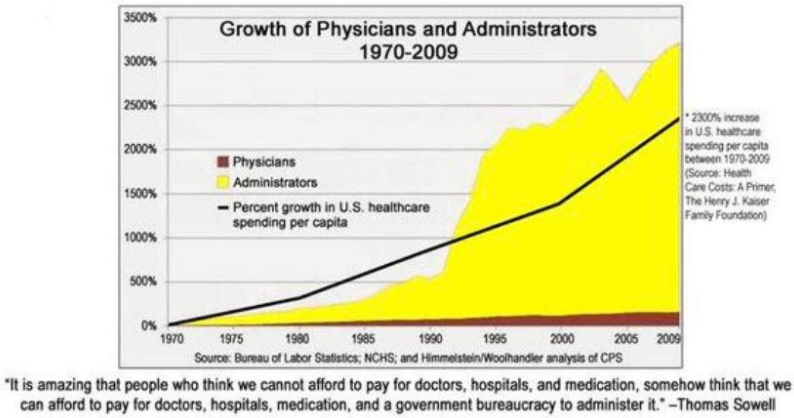
```
p <- d  %>%
  ggplot(aes(x = mid_quarter_month, y = total_hours / 1000, fill =
occupation)) +
  scale_fill_manual(values = occ_palette) +
  theme(legend.position = "right") +
  labs(caption = the_caption,
       x = str_wrap("Health care professionals who work in other
industries adjusted for by
                    subtracting around 14% over all years, based on a
rough estimate from
                    the ABS Census of Population and Housing 2011.",
120),
       fill = "Occupation",
       subtitle = "Hours by occupation of workers in Australia's
health care and social assistance industry") +
  theme(axis.title.x = element_text(size = 9, hjust = 0, colour =
"grey"))

# chart: stacked so we see absolute size
p + geom_area(position = "stack") +
  scale_y_continuous(label = comma) +
  labs(y = "Millions of hours worked per quarter",
       title = "Steady growth over time")

# chart: filled to top, showing proportions
p + geom_area(position = "fill") +
  scale_y_continuous(label = percent_format(accuracy = 1)) +
  labs(y = "Proportion of hours worked",
       title = "More community and personal service workers and less
labourers")


#==================Medical practitioners=======
# Out of curiousity, let's look more at the breakdown of those health
professionals

med_prac <- c(
  "2530 Medical Practitioners nfd"
,
  "2531 General Practitioners and Resident Medical Officers"
,
  "2532 Anaesthetists"
,
  "2533 Specialist Physicians"
,
  "2534 Psychiatrists"
,
  "2535 Surgeons"
,
  "2539 Other Medical Practitioners"
)


profs_only <- eq08 %>%
  filter(occupation4 %in% health_profs) %>%
  mutate(med_prac = if_else(occupation4 %in% med_prac, "Medical
practitioner", "Other health professional")) %>%
```

```
  group_by(mid_quarter_month, med_prac) %>%
  summarise(total_hours = sum(total_hours))

# chart: medical practitioners as a proportion of health professionals
profs_only %>%
  ggplot(aes(x = mid_quarter_month, y = total_hours, fill = med_prac))
+
  geom_area(position = "fill") +
  scale_y_continuous(label = percent) +
  labs(x = str_wrap("Medical practitioners defined as GPs, Resident
Medical Officers, Anaesthetists, Specialist Physicians,
       Psychiatrists, Surgeons, and other Medical Practitioners.
'Other health professional' examples includes nurses,
       pharmacists, midwives, nutritional practitioners, dental
practitioners.", 120),
       fill = "",
       y = "Percentage of all health professionals' hours",
       subtitle = "Hours worked by all health professionals (unit
group code 2500 to 2544)",
       title = "Medical practitioners' labour has remained a constant
proportion of health professionals'",
       caption = the_caption) +
  theme(axis.title.x = element_text(size = 9, hjust = 0, colour =
"grey"))
#=================which are growing fast?============

# chart: lollipop of fastest growing or shrinking occupations
eq08 %>%
  # remove any 'not further defined' residual categories so we can
focus on real occupations
  filter(!grepl("nfd$", occupation4)) %>%
  mutate(yr = year(mid_quarter_month)) %>%
  filter(yr %in% c(1987, 2019)) %>%
  group_by(occupation4, yr) %>%
  summarise(total_hours = sum(total_hours)) %>%
  group_by(occupation4) %>%
  arrange(yr) %>%
  mutate(start_hours = total_hours[1],
         growth = total_hours / start_hours - 1,
         growth_backwards = start_hours / total_hours - 1,
         growth_either = ifelse(growth > 0, growth,
-growth_backwards)) %>%
  filter(yr == max(yr)) %>%
  ungroup() %>%
  arrange(desc(abs(growth_either))) %>%
  slice(1:25) %>%
  mutate(occupation4 = fct_reorder(str_sub(occupation4, 6),
growth_either)) %>%
  ggplot(aes(y = occupation4, yend = occupation4, x = growth_either,
xend = 0)) +
  geom_segment(size = 2, aes(colour = growth_either)) +
  geom_point(aes(size = total_hours / 1000)) +
  scale_x_continuous(label = percent) +
  scale_size_area(label = comma_format(accuracy = 1)) +
  scale_colour_viridis_c(option = "C", label = percent, guide =
'none') +
  labs(x = "Growth in employment hours from 1987 to 2019 (if
```

```
increasing),\nor from 2019 to 1987 (if decreasing)",
       y = "",
       caption = the_caption,
       colour = "",
       size = "Million hours per year in 2019",
       title = "The biggest growing and shrinking occupations by
employment hours",
       subtitle = paste0("Two occupations grew by 3,000% from nearly
non-existent in 1987.\n",
                         "Textile footwear machine operators' hours
were 1,500% more in 1987 than (near zero) in 2019."))
```

Other supplementary material

Here's the original image that prompted me to think about all this:



And here's a screenshot from the ABS Census tablebuilder, for anyone curious about which industries other than Health Care and Social Assistance employ medical professionals:

<< OCCP - 2 Digit Level	Health Professionals	Total
INDP - 1 Digit Level		
Agriculture, Forestry and Fishing	143	143
Mining	2,216	2,216
Manufacturing	3,228	3,228
Electricity, Gas, Water and Waste Services	730	730
Construction	2,543	2,543
Wholesale Trade	994	994
Retail Trade	16,182	16,182
Accommodation and Food Services	468	468
Transport, Postal and Warehousing	1,173	1,173
Information Media and Telecommunications	154	154
Financial and Insurance Services	1,644	1,644
Rental, Hiring and Real Estate Services	154	154
Professional, Scientific and Technical Services	3,190	3,190
Administrative and Support Services	3,834	3,834
Public Administration and Safety	14,116	14,116
Education and Training	4,989	4,989
Health Care and Social Assistance	373,609	373,609
Arts and Recreation Services	346	346
Other Services	1,632	1,632
Inadequately described	1,375	1,375
Not stated	1,018	1,018
Not applicable	0	0
Total	433,726	433,726

Data Source : Census of Population and Housing, 2011, TableBuilder