Final Portfolio

Kai Shuen Neo

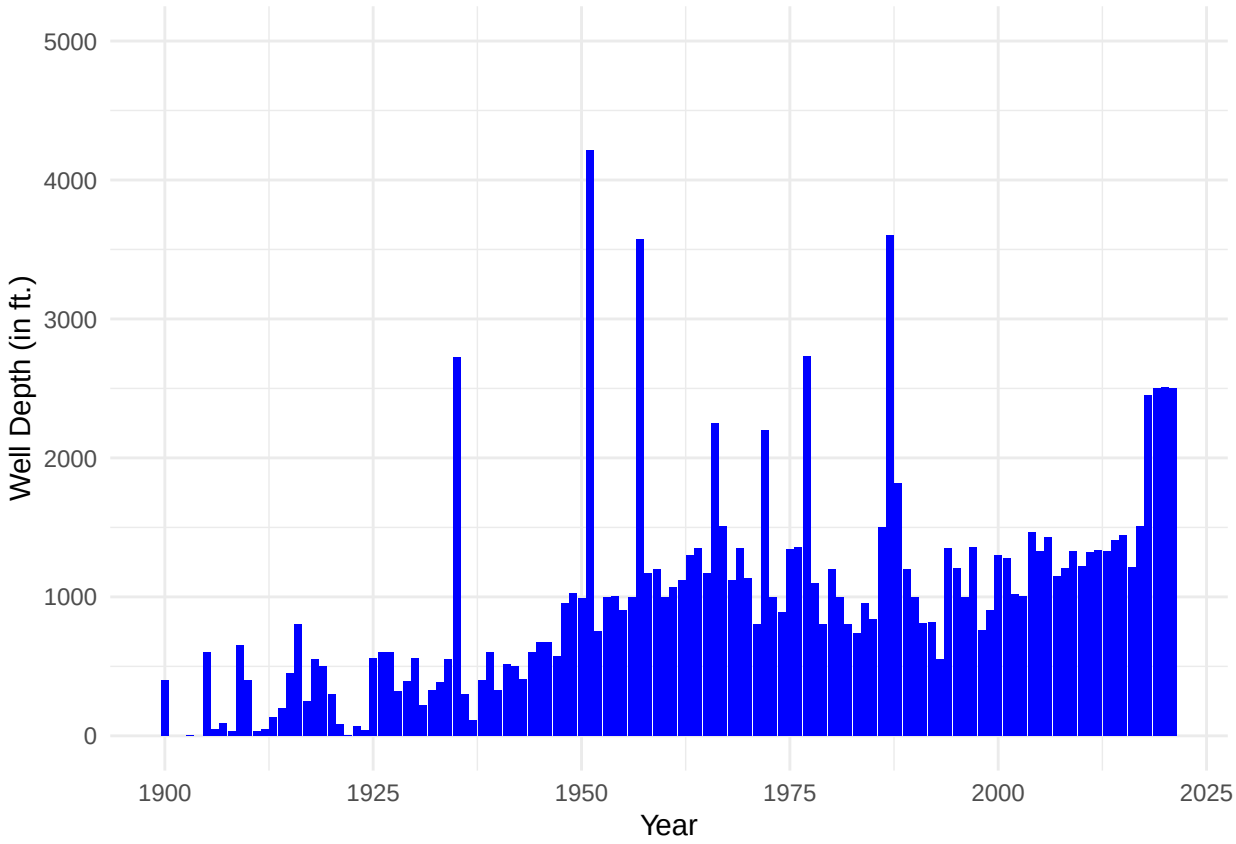
2024-02-27

Table 1: Summary Statistics on Arizona Wells (in ft.)

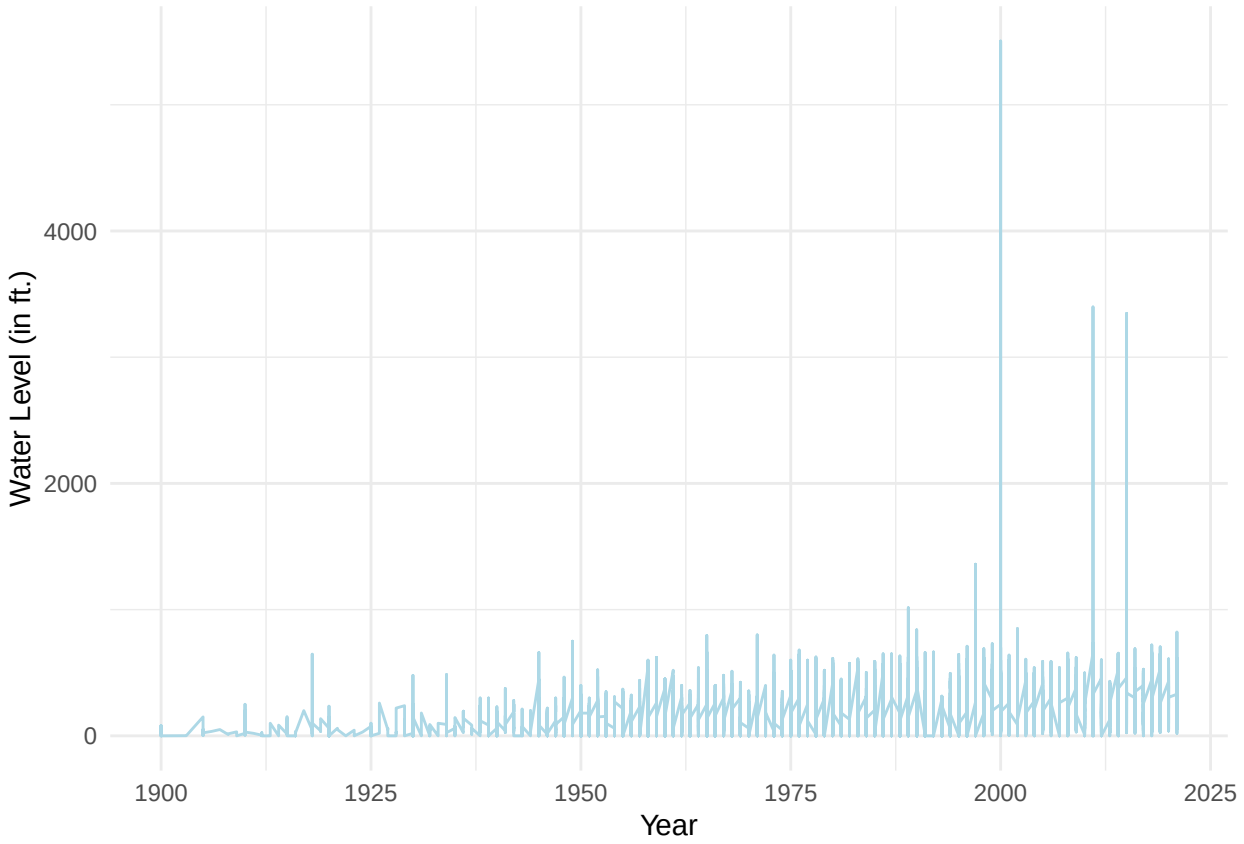
Basin	Total Well Depth	Mean Well Depth	Median Well Depth	Total Water Level	Mean Water Level	Median Water Level	Total Casing Depth	Mean Casing Depth	Median Casing Depth
Douglas	162366	297.9193	250	64310	118.0000	65	123867	227.2789	200
Douglas INA	608225	308.9005	250	246866	125.3763	100	558070	283.4281	230
Willcox	1959587	395.7163	328	891087	179.9449	145	1817025	366.9275	300

Table 1 - A table with the summary statistics on key characteristics for Arizona Wells. Measurements (in ft.) were taken in 3 Basins (Douglas INA, Douglas, Willcox) in drought-stricken Southeast Arizona.

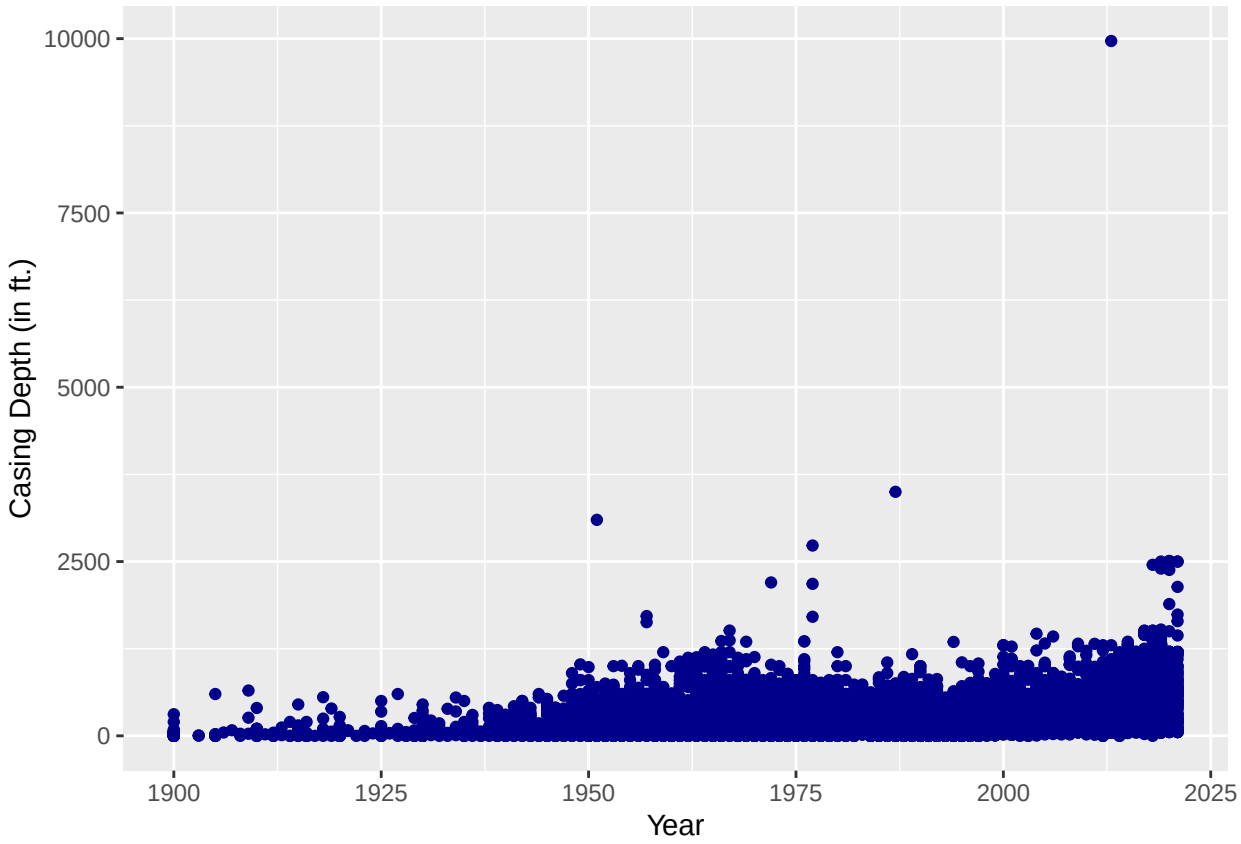
Table Notes: Well depth refers to the depth that the well was dug to. Water level refers to how deep one has to go to get water. Casing depth refers to the depth of the solid casing that was lining the well.



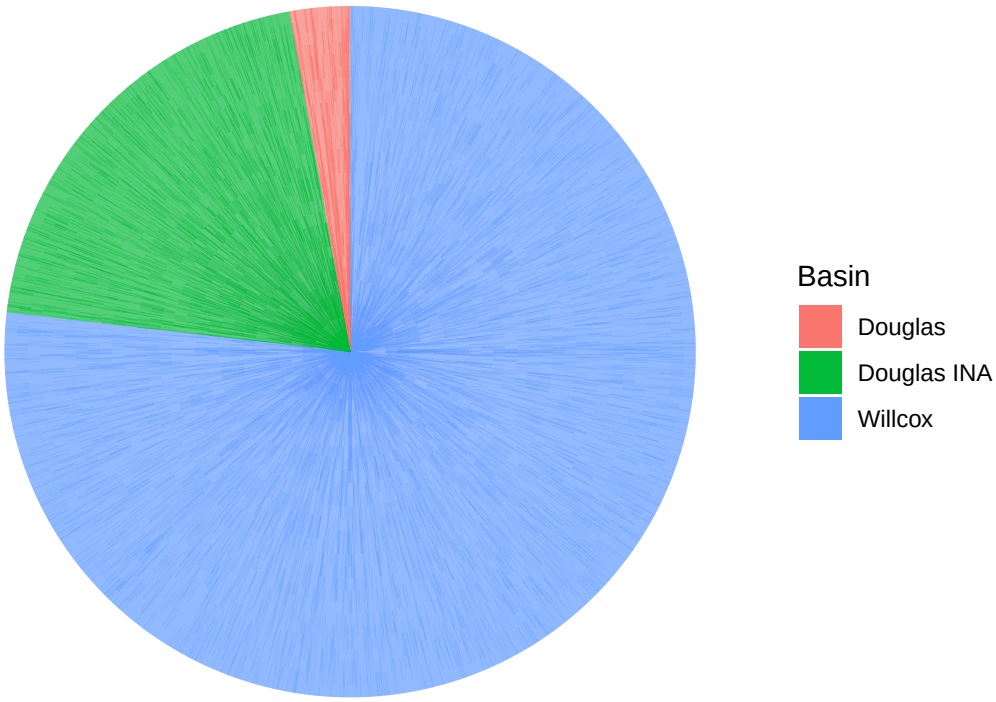
Plot 2 - This bar plot describes the change in Well Depth (in ft.) over the years. As well depth refers to the depth that the well was dug to, it is observed that irregardless of basins, there is an overall gradual increase in the depth of wells from 1900 to 2021, despite several anomalies whereby there were peaks in well depth for certain years. This tells us that wells are getting dug deeper and deeper over the years, from a shallow well of close to 0 ft. in the 1900s, to a very deep well of at least 2500 ft.



Plot 3 - This line plot describes Water Level (in ft.) over the years. Water level refers to how deep one has to go to get water. Water level is actually considerably constant and hovering at about less than 1000 ft. over the years, except for year 2000. This means that the surface level in which one has to reach for, did not change much over the years despite the changes in the depth of wells. Perhaps the anomaly in year 2000 might be due to certain natural conditions (e.g. extreme drought), which might have caused surface level of water in the wells to drop significantly. As such, people had to reach further to obtain water due to the drop in water quantity.



Plot 4 - This scatter plot describes the Casing Depth (in ft.) over the years. Casing depth refers to the depth of the solid casing that was lining the well. Overall, casing depth was increased only mildly gradually over the years. This meant that the depth of the solid casing lining the well was increasingly getting thicker. There were several outliers in the plot, with significantly high casing depth. However, these should not affect the overall observation since it is very insignificant in calculation. This could signify that there wasn't exactly a need for thicker casing depth to be built to be lining the well since drought conditions might still be ever present.



Plot 5 - This pie chart describes the distribution of basins in Southeast Arizona. There are namely 3 basins in this dataset, namely Douglas, Douglas INA and Willcox. Willcox makes up the largest distribution of basin in this dataset for Arizona Wells, followed by Douglas INA and lastly, Douglas.

Code Output

```
# import libraries
library(dplyr)
library(ggplot2)
library(tidyverse)
library(lubridate)
library(kableExtra)

# import data set
arizona_wells <- read.csv("Wells By Basin.csv")

# understanding data
unique(arizona_wells$Basin)
unique(arizona_wells$Well.Type)
unique(arizona_wells$Casing.Depth..ft.)
unique(arizona_wells$Water.Level..ft.)
unique(arizona_wells$Well.Depth..ft.)


# data quality
date <-  as.Date(arizona_wells$Drill.Date,"%Y-%m-%d %H:%M:%S")
year <- as.numeric(format(date,'%Y'))

arizona_wells$year <- year

aw <- arizona_wells %>%
  filter(!is.na(Basin)) %>%
  filter(!is.na(Well.Type)) %>%
  filter(!is.na(Well.Depth..ft.)) %>%
  filter(!is.na(Casing.Depth..ft.)) %>%
  filter(!is.na(Drill.Date)) %>%
  filter(!is.na(Application.Date)) %>%
  filter(!is.na(Water.Level..ft.)) %>%
  filter(!is.na(year))

# table - summary statistic of the well depth, water level, casing depth
water_level_by_year <- aw %>%
  group_by(Basin) %>%
  summarize(total_well_depth=sum(Well.Depth..ft.),
            mean_well_depth=mean(Well.Depth..ft.),
            median_well_depth=median(Well.Depth..ft.),
            total_water_level=sum(Water.Level..ft.),
            mean_water_level=mean(Water.Level..ft.),
            median_water_level=median(Water.Level..ft.),
            total_casing_depth=sum(Casing.Depth..ft.),
            mean_casing_depth=mean(Casing.Depth..ft.),
            median_casing_depth=median(Casing.Depth..ft.))

knitr::kable(
  water_level_by_year,
  row.names = NA,
  col.names = c('Basin',
                'Total Well Depth',
```

```
                'Mean Well Depth',
                'Median Well Depth',
                'Total Water Level',
                'Mean Water Level',
                'Median Water Level',
                'Total Casing Depth',
                'Mean Casing Depth',
                'Median Casing Depth'),
  caption = "Summary Statistics about Arizona Wells (in ft.)"
)


# bar plot
bar_plot <- ggplot(aw, aes(x=year, y=Well.Depth..ft.)) +
  geom_bar(stat="identity", position="dodge", fill="blue") +
  labs(x="Year", y="Well Depth (in ft.)") +
  expand_limits(y=c(0, 5000)) +
  theme_minimal()

print(bar_plot)

# line plot
line_plot <- ggplot(aw, aes(x=year, y=Water.Level..ft.)) +
  geom_line(stat="identity", position="dodge", color="lightblue") +
  labs(x="Year", y="Water Level (in ft.)") +
  theme_minimal()

print(line_plot)

# scatter plot
scatter_plot <- ggplot(aw, aes(x=year, y=Casing.Depth..ft.)) +
  geom_point(color="darkblue") +
  labs(x="Year", y="Casing Depth (in ft.)")

print(scatter_plot)

# pie chart
pie <- ggplot(aw, aes(x="", y= Basin, fill=Basin)) +
  geom_bar(stat = "identity", width = 1)+
  coord_polar("y", start = 0) +
  theme_void()
print(pie)
```