

# Rapport du Projet

# Visualisation de données

« Élections USA 2020 »



Suivi par: Mme BOUHAFS Asma

NOM : TACHOUR Prénom : Omar

N° étudiant: 16706265

2020/2021

# **Sommaire:**

1. Description des bases de données utilisées4
2. Introduction5
3. Les élections présidentielles5
Le total des votes obtenu par chaque candidat6
Comparaison entre les candidats en termes du nombre de comtés gagné: 7
❖ Nombre de comtés où chaque parti politique (DEM, REP, LIB) est gagnant ou perdant:
Cartographie des comtés où les deux candidats (Trump et Biden)
ont été majoritaires:9
❖ Nombre de comtés par sexe
➤ Nombre de comtés où female > male
Nombre de comtés où female ≤ male
Comparaison entre le nombre de femmes et le nombre d' hommes
dans chaque état:10
Nombre de comtés gagné par chaque candidat dans 2 cas:
➤ Cas 1 : female > male
$ ightharpoonup$ Cas 2 : female $\leq$ male :
Total votes des deux candidats (Biden et Trump) par sexe et
par candidats:12
❖ Total votes dans tous les états par candidat :13
❖ L'âge moyen des votants pour les deux candidats (Joe Biden et Donald Trump) par états :
❖ L' âge moyen dans les états où female_percentage > 50
<b>♦</b> L' âge moyen dans les états où female percentage ≤ 5015
❖ L'âge moyen des votants dans chaque état par candidat :16
❖ Âge moyen des votants dans les états où Biden est gagnant et
dans les états où Trump est gagant :
❖ Total de votes pour les 2 candidats par sexe et âge moyen :

➤ Âge moyen > 40 ans
$ ightharpoonup \hat{A}$ ge moyen $\leq 40$ ans
Étude des élections par le sexe et le pourcentage des femmes :
➤ Pourcentage de femmes dans chaque état20
➤ Pourcentage de femmes des votants pour Joe Biden par états
➤ Pourcentage de femmes des votants pour Donald Trump par états
<u>Étude des élections par la population</u> 23
❖ La population de chaque état23
❖ Population de chaque état où (Biden ou Trump) est gagnant23
Population par sex (masculine et féminine) dans chaque état où
(Biden ou Trump) est gagnant:24
<b>❖</b> Corrélation entre Population et candidat25
<b>❖</b> Carte de corrélation26
❖ Corrélation entre candidat, âge moyen, female pourcentage27
❖ Corrélation entre female, male, population, total vote28
<b>❖</b> Taux de participation par états29
<b>❖</b> Population éligible au vote29
<b>❖</b> Population en âge de voter par états30
Nombre de votants en prison par états30
Nombre de votants en probation par État31
❖ Pourcentage des citoyens non américains votants par états31
Nombre de votants en libération conditionnelle par État32
Nombre de citoyens ineligibles au vote par états32
4. Conclusion :

# 1. Description des bases de données utilisées:

• https://www.kaggle.com/unanimad/us-election-2020

	state	county	candidate	party	total_votes	won
0	Delaware	Kent County	Joe Biden	DEM	44552	True
1	Delaware	Kent County	Donald Trump	REP	41009	False
2	Delaware	Kent County	Jo Jorgensen	LIB	1044	False
3	Delaware	Kent County	Howie Hawkins	GRN	420	False
4	Delaware	New Castle County	Joe Biden	DEM	195034	True
5	Delaware	New Castle County	Donald Trump	REP	88364	False
6	Delaware	New Castle County	Jo Jorgensen	LIB	2953	False
7	Delaware	New Castle County	Howie Hawkins	GRN	1282	False
8	Delaware	Sussex County	Donald Trump	REP	71230	True
9	Delaware	Sussex County	Joe Biden	DEM	56682	False

Cette base de données calcule le total votes par (état, département, candidats, parti) et indique si le candidat à gagné dans ce département.

#### • https://www.kaggle.com/headsortails/covid19-us-county-jhu-data-demographics

	fips	county	state	state_code	male	female	median_age	population	female_percentage	lat	long
0	1001	Autauga County	Alabama	AL	26874	28326	37.8	55200	51.315217	32.534923	-86.642730
1	1003	Baldwin County	Alabama	AL	101188	106919	42.8	208107	51.376936	30.727479	-87.722564
2	1005	Barbour County	Alabama	AL	13697	12085	39.9	25782	46.873788	31.869581	-85.393210
3	1007	Bibb County	Alabama	AL	12152	10375	39.9	22527	46.055844	32.998628	-87.126475
4	1009	Blount County	Alabama	AL	28434	29211	40.8	57645	50.673953	33.980869	-86.567380
5	1011	<b>Bullock County</b>	Alabama	AL	5663	4689	39.6	10352	45.295595	32.100525	-85.715679
6	1013	Butler County	Alabama	AL	9368	10657	40.7	20025	53.218477	31.752429	-86.680276
7	1015	Calhoun County	Alabama	AL	55315	59783	39.7	115098	51.940955	33.771416	-85.826039
8	1017	Chambers County	Alabama	AL	16257	17569	43.0	33826	51.939337	32.914345	-85.392030
9	1019	Cherokee County	Alabama	AL	12820	13033	45.9	25853	50.411944	34.175905	-85.603801

Cette base de données contient les informations démographiques (male, female, median\_age, population, female percentage) au niveau des comtés américains basées sur la version (la plus récente) 2014-2018..

https://www.kaggle.com/imoore/2020-us-general-election-turnout-rates.

State	Source	Official/Unofficial	Total Ballots Counted (Estimate)	Vote for Highest Office (President)	VEP Turnout Rate	Voting- Eligible Population (VEP)	Voting-Age Population (VAP)	% Non- citizen	Parole	Total Ineligible Felon	Overseas Eligible	State Abv	
United States	NaN	NaN	158,835,004	NaN	66.4%	239,247,182	257,605,088	7.8%	616,440	3,294,457	4,971,025	NaN	
Alabama	https://www2.alabamavotes.gov/electionnight/st	Unofficial	2,306,587	2,297,295	62.6%	3,683,055	3,837,540	2.3%	10,266	67,782	NaN	AL	
Alaska	https://www.elections.alaska.gov/results/20 GEN	NaN	367,000	NaN	69.8%	525,568	551,117	3.4%	1,348	6,927	NaN	AK	
Arizona	https://results.arizona.vote/#/featured/18/0	NaN	3,400,000	NaN	65.5%	5,189,000	5,798,473	8.9%	7,536	93,699	NaN	AZ	
Arkansas	https://results.enr.clarityelections.com/AR/10	Unofficial	1,212,030	1,206,697	55.5%	2,182,375	2,331,171	3.6%	24,698	64,974	NaN	AR	
California	https://electionresults.sos.ca.gov/	Unofficial	16,800,000	NaN	64.7%	25,962,648	30,783,255	15.0%	102,586	207,316	NaN	CA	
Colorado	https://results.enr.clarityelections.com/CO/10	NaN	3,295,000	NaN	76.4%	4,313,054	4,595,504	5.7%	0	18,905	NaN	со	ı
Connecticut	NaN	NaN	1,850,000	NaN	71.1%	2,603,327	2,839,560	7.7%	4,452	17,720	NaN	СТ	ı
Delaware	https://elections.delaware.gov/results/html/in	Unofficial	507,805	502,392	70.5%	720,531	780,822	5.8%	350	15,013	NaN	DE	

Cette base de données contient des informations sur le taux de participation est le pourcentage d'électeurs éligibles qui ont voté lors d'une élection. L'éligibilité varie selon les pays et la population éligible au vote ne doit pas être confondue avec la population adulte totale. L'âge et le statut de citoyen font souvent partie des critères utilisés pour déterminer l'éligibilité, ainsi que la population votante depuis la prison, par parole et en probation.

#### 2. Introduction

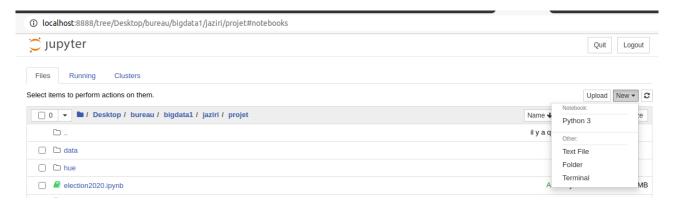
- ➤ Dans ce projet de visualisation de données, 3 bases de données qui contient des informations sur les élections américaines 2020 seront traitées, elles vont être récupérées à partir du site « kaggle », le lien ci-dessus, ce dernier permet de collecter et publier plusieurs ensembles de données sur n'importe quel sujet.
- ➤ Mon travail consiste à réaliser une analyse compète et profonde sur les élections présidentielles 2020 aux états unis d'amérique (comparaison entre les résultats de Donald Trump et Joe Biden et études de taux de participation en différents termes).
- Pour cela je vais utiliser 3 bases de données:
  - "president\_county\_candidate.csv"
  - o "president counties.csv"
  - "taux participation.csv"
- ➤ Pour les outils de la visualisation de données, j'ai utilisé des frameworks Python (Plotly, matplotlib, seaborn, Pandas, Numpy, scipy, sklearn).

# 3. Les élections présidentielles :

Dans cette partie j'ai créé un notebook sur jupyter notebook que je lance depuis le terminal

```
achour@tachour-HP-Pavilion-Laptop-14-ce0xxx:~$ jupyter-notebook
   19:15:31.226 NotebookApp] Serving notebooks from local directory: /home/tacho
   19:15:31.227 NotebookApp] Jupyter Notebook 6.1.6 is running at:
   19:15:31.227 NotebookApp] http://localhost:8888/?token=12e9994d0bb8e6372d3e44
63e7e074b44cc8d1373cf9c428
                              or http://127.0.0.1:8888/?token=12e9994d0bb8e6372d
3e4463e7e074b44cc8d1373cf9c428
            227 NotebookApp] Use Control-C to stop this server and shut down all
 kernels (twice to skip confirmation).
    To access the notebook, open this file in a browser:
        file:///home/tachour/.local/share/jupyter/runtime/nbserver-4087-open.htm
    Or copy and paste one of these URLs:
        http://localhost:8888/?token=12e9994d0bb8e6372d3e4463e<u>7e074b44cc8d1373cf</u>
9c428
    or http://127.0.0.1:8888/?token=12e9994d0bb8e6372d3e4463e7e074b44cc8d1373cf
9c428
   19:16:03.263 NotebookApp] Kernel started: 89a3b6c0-b876-41f4-beee-e7567960315
  name: python3
                NotebookApp] 404 GET /static/plotly.js?v=20210102191530 (127.0.0
   18.560000ms referer=http://localhost:8888/notebooks/Desktop/bureau/bigdata1/
```

Ensuite on clique sur new pour créer un fichier python :



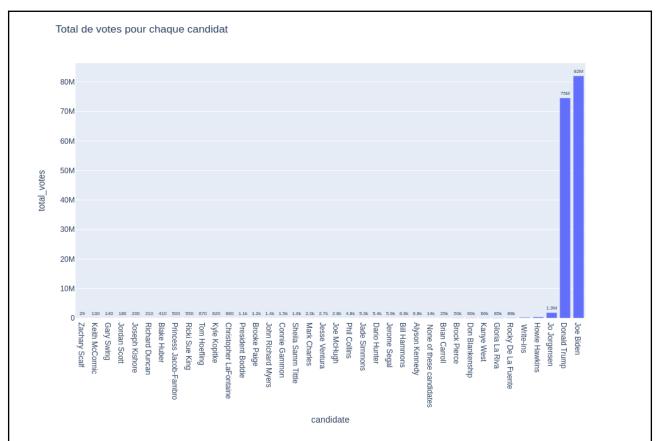
On commence toujours par l'installation de quelques framworks python et l'importation des bibliothèques.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

import plotly.graph_objects as go
import plotly.io as pio
import plotly.express as px
import plotly.express as ex
import plotly.express as ex
import plotly.express as py
import plotly.offline as pyo
from plotly.subplots import make_subplots

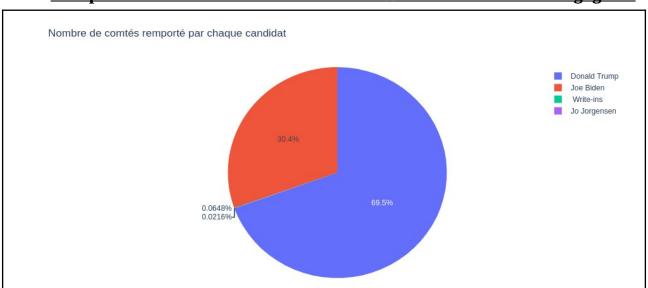
import seaborn as sns
from scipy import stats
from sklearn.preprocessing import LabelEncoder
encoder = LabelEncoder()
```

### **Le total de votes obtenu par chaque candidat**



Ce graphique illustre le classement de chaque candidat dans les agrégateurs de sondages de septembre 2019 jusqu'à la fin des élections. L'ancien vice-président Joe Biden, le candidat démocrate, émarge à 82 millions, affichant une supériorité de 7 millions de votes par rapport au président sortant Donald Trump, le candidat républicain.

#### Comparaison entre les candidats en termes du nombre de comtés gagnées

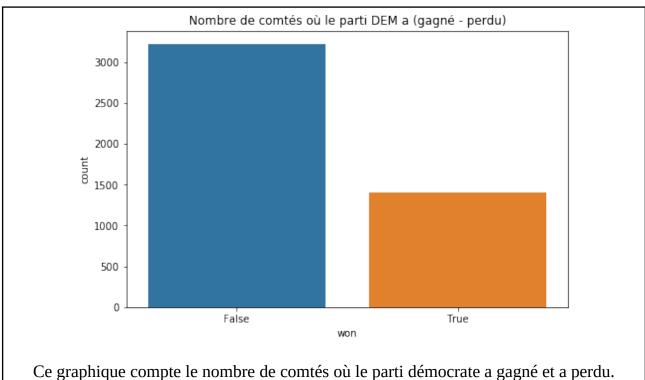


Ce diagramme circulaire définit le nombre de comtés où chaque candidat a gagné, l'exprésident Donald Trump a été en tête dans 69,5% des départements en USA, suivi par Joe Biden avec un pourcentage de 30,4%. Ainsi que les candidats Write-ins et Jo Jorgensen

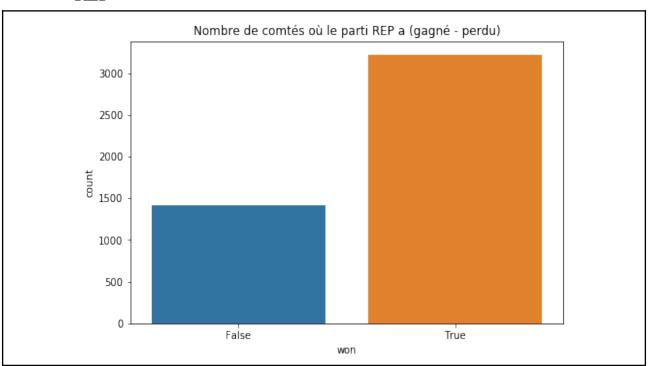
qui ne dépassent pas tous les deux 0,1%.

# Nombre de comtés où chaque parti politique (DEM, REP, LIB) est gagnant ou perdant:

#### **>** DEM

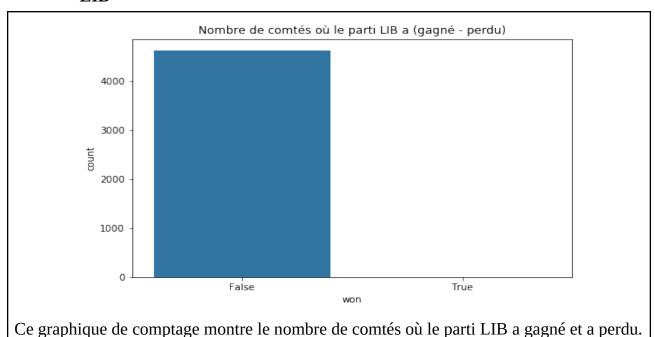


#### **≻** REP

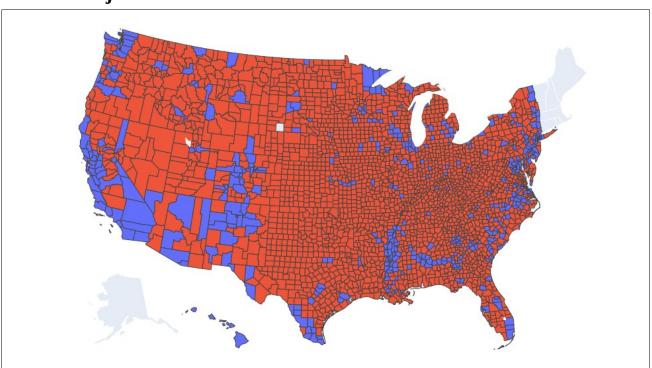


Ce graphique compte le nombre de comtés où le parti républicain a gagné et a perdu.

#### **≻** LIB



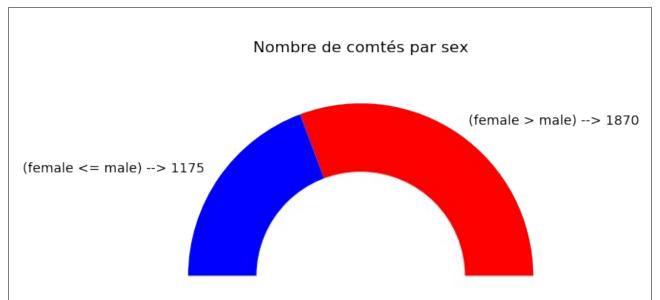
# **Cartographie des comtés où les deux candidats (Trump et Biden) ont été majoritaires:**



Cette carte montre clairement que le candidat Donald Trump est majoritaire en terme de nombre de départements gagnés. Malgré ça, Joe Biden a remporté les élections présidentielles.

### **♦** Nombre de comtés par sex :

- > Nombre de comtés où female > male
- > Nombre de comtés où female ≤ male\_

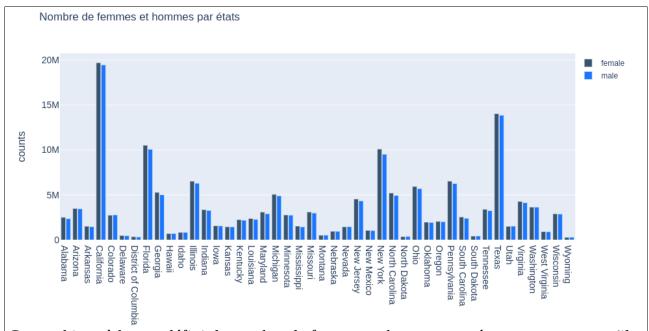


Ce graphique circulaire compte le nombre de comtés par sex :

- Cas 1: le nombre des femmes > au nombre des hommes
- Cas 2: le nombre des femmes < au nombre des hommes

On constate que le nombre de comtés où les femmes ont voté plus que les hommes est supérieur à celui où les hommes ont voté plus.

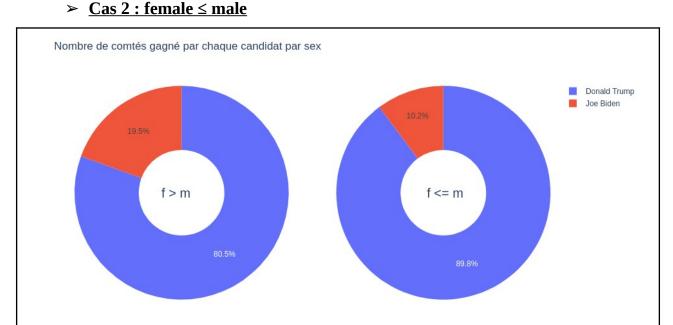
**Comparaison entre le nombre de femmes et le nombre d' hommes dans chaque état:** 



Ce graphique à barres définit le nombre de femmes et hommes par états, on constate qu'il y a une toute petite différence entre les deux valeurs presque dans chaque états (femmes > hommes).

# Nombre de comtés gagné par chaque candidat dans 2 cas:

➤ Cas 1 : female > male

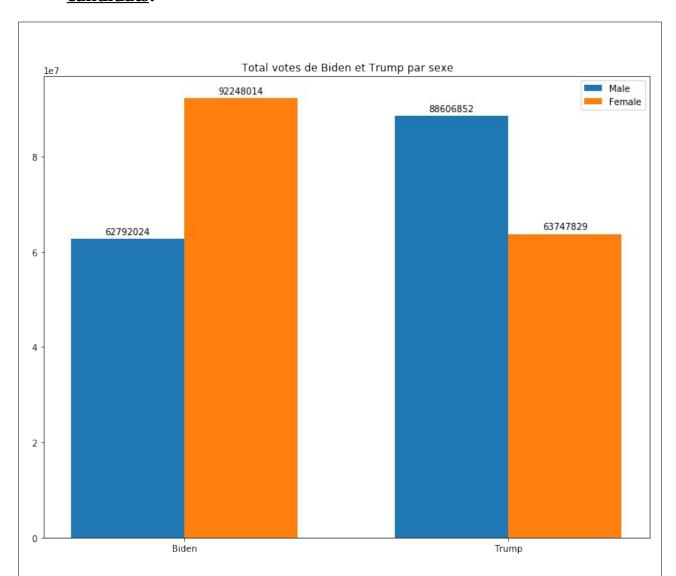


Ces deux diagrammes circulaires définissent le nombre de comtés remportés par chaque candidat dans 2 cas:

- Cas1 : quand le nombre de femmes > hommes
- Cas2 : quand le nombre de femmes <= hommes

On voit clairement que le nombre de comtés gagné par Donald Trump est supérieur à celui de Joe Biden dans les deux cas.

# **❖** Total votes pour les deux candidats (Biden et Trump) par sexe et par candidats:

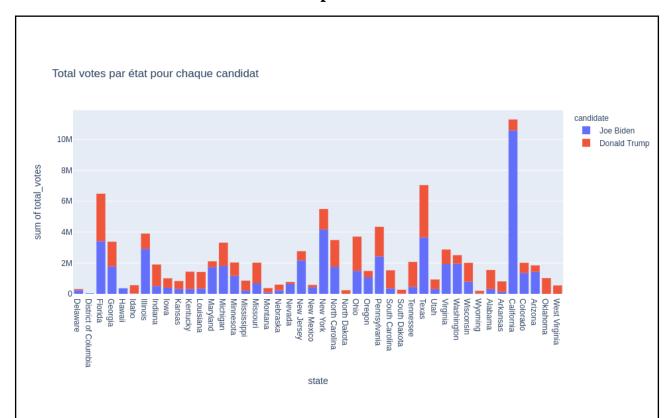


Cette figure fait la comparaison entre la population par sexe(male/female) dans les comtés gagnées par Biden et Trump.

D'après la figure on voit que même si Donald trump a remporté un nombre de comtés incomparable à celui de Biden mais ce dernier le dépasse en terme de population :

Total population de comtés où Biden est gagnant : 155040038 Total population de comtés où Trump est gagnant : 152354681 D'après la figure on observe que la population dans les département de Biden comporte plus de femmes que les homme, c'est le cas contraire pour Trump.

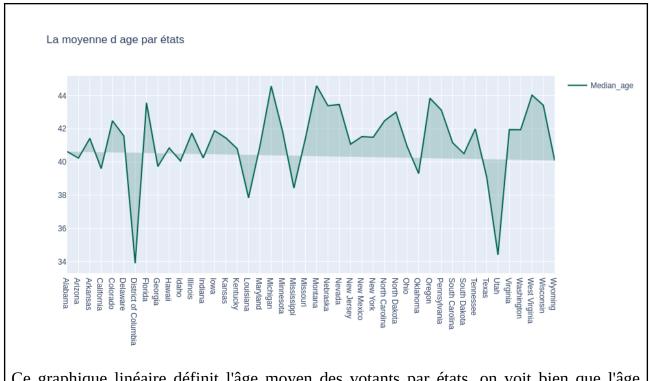
# **Total votes dans tous les états par candidat :**



Cet histogramme illustre une comparaison entre le total de votes pour chaque candidat par état.

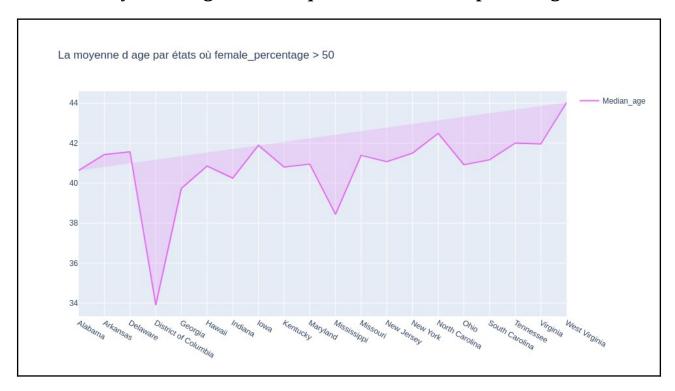
On constate que les deux côtés sont majoritaires dans le même nombre d'états, par contre Joe Biden est souvent le gagnant dans les États où il y a beaucoup de votes comme California, New York, Florida, Texas...

# **L'âge moyen des votants pour les deux candidats (Joe Biden et Donald Trump) par états :**



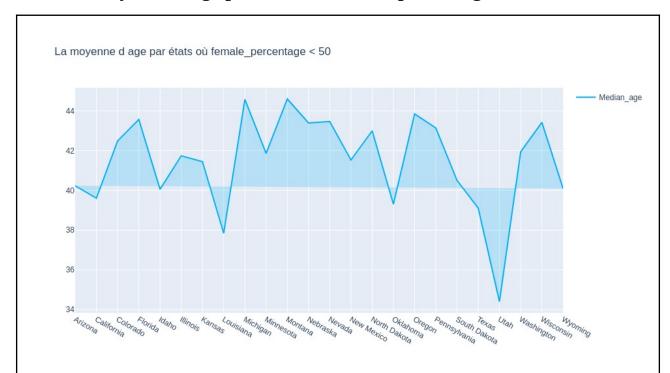
Ce graphique linéaire définit l'âge moyen des votants par états, on voit bien que l'âge moyen est situé entre (34 et 44 ans) d'une manière globale, mais la plupart des villes ont un âge moyen situé entre (40 et 44 ans).

### **❖** La moyenne d'âge dans chaque états où female percentage > 50



Ce graphique linéaire définit l'âge moyen des votants par états, dans le cas où le pourcentage de femmes est supérieur à 50%, on constate que l'âge moyen est situé entre 40 et 44 ans, sauf dans l'état "District of Columbia".

# **❖** La moyenne d'âge par états où female percentage <= 50

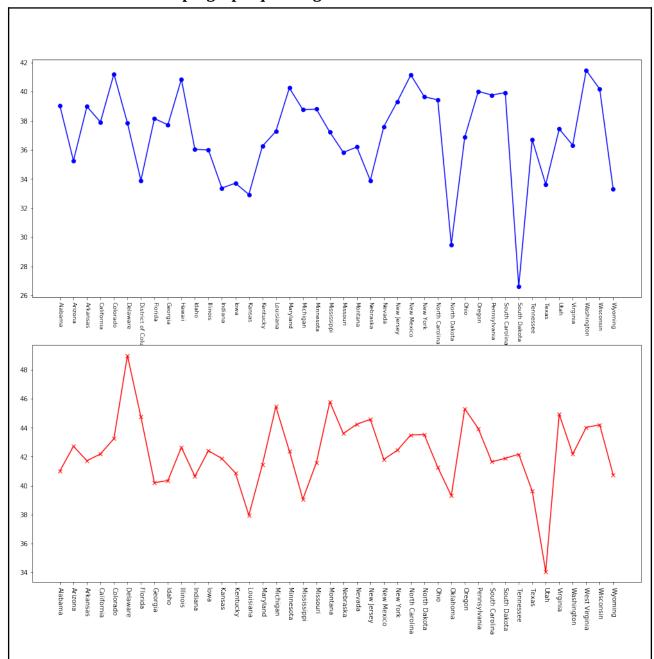


Ce graphique linéaire définit l'âge moyen des votants par états, dans le cas où le pourcentage de femmes est inférieur à 50%, on constate que l'âge moyen est situé entre 40 et 44 ans, sauf dans l'état Utah.

# **❖** L'âge moyen des votants dans chaque état par candidat :

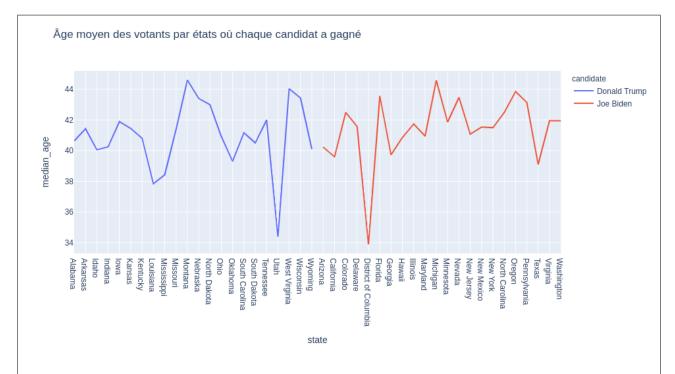
> Joe Biden : graphique bleu.

> Donald Trump: graphique rouge.



Ces deux histogrammes représentent l'intervalle de la moyenne d'âge des votants pour les deux candidats par états. On authentifie que la moyenne d'âge des habitants des états de Joe Biden est [26 - 42] et [34 - 48] pour Donald Trump. Par conséquent, la population de J.Biden est plus jeune que celle de Trump.

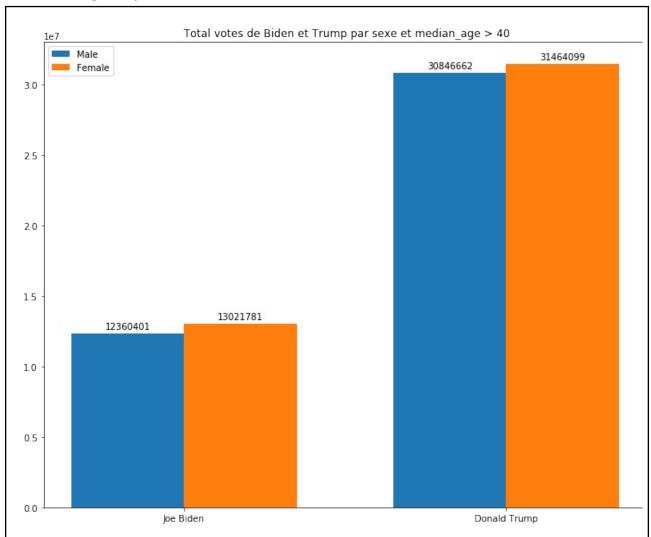
# \* Âge moyen des votants dans les états où Biden est gagnant et dans les états où Trump est gagant :



Ces deux diagrammes linéaires montrent l'intervalle de l'âge moyen par états où chaque candidat a gagné, on constate que les résultats des deux candidats sont presque équivalents en termes des états gagnés et en termes de l'âge moyen.

# **❖** Total de votes pour les 2 candidats par sex et âge moyen.

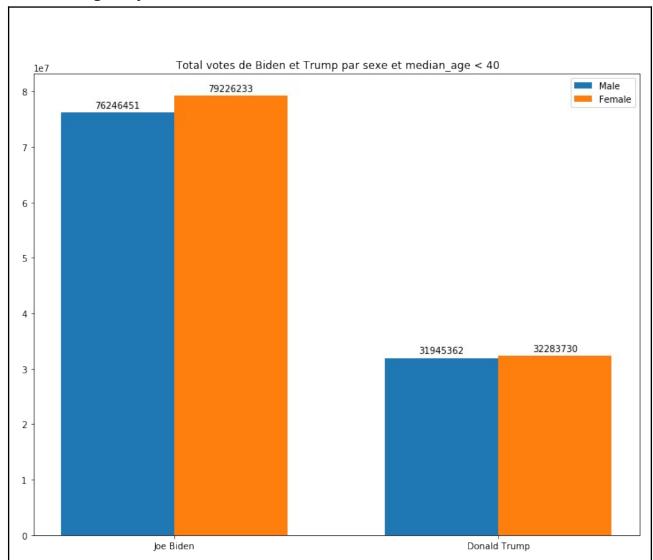
#### **>** Âge moyen > 40



Ces deux graphiques à barres montrent la population votante pour les deux candidats ayant un âge moyen supérieur à 40 ans, par sex.

On voit que Trump est majoritaire avec un nombre (31 millions hommes / 31 millions femmes). De l'autre côté, Joe Biden remporte (13 millions femmes et 12 millions hommes) ce qui fait le tiers de Trump.

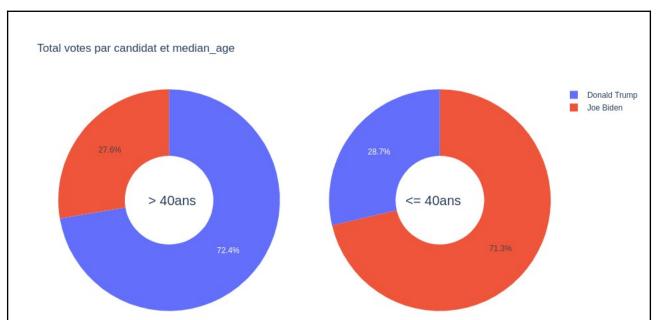
#### ➤ Âge moyen < 40



C'est deux graphiques à barres représentant la population votante pour les deux candidats ayant un âge moyen > à 40 ans, par sex.

On voit que Trump a remporté un nombre de (32 millions hommes / 32 millions femmes). De l'autre coté, Biden remporte un chiffre très important qui vaut presque le triple de Trump, (76 millions hommes et 79 millions femmes).

D'après ces deux derniers graphiques à barres, on comprend que la population de Joe Biden est plus jeunes que celle de Trump, d'une autre manière, la population de Trump est vielle par rapport à celle de Joe Biden.

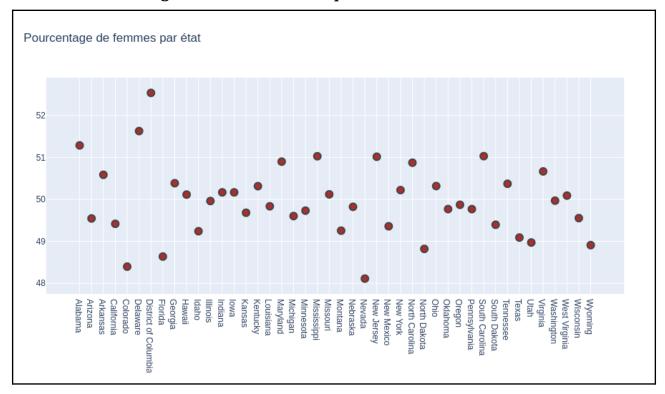


Ces deux diagrammes circulaires font une comparaison entre le total de votes entre deux candidats en fonction de l'âge moyen (âge moyen > 40 ans, âge moyen <= 40 ans)

On remarque clairement que Donald Trump a remporté 72.4% des votes dans le cas où l'âge moyen est supérieur à 40 ans, contre 27.6% des voix pour Joe Biden. En revanche, ce dernier a gagné 71.3% des votes contre 28.7% pour Trump dans le cas où les votants ont un âge moyen <= à 40 ans.

# **Étude des élections par le pourcentage des femmes**

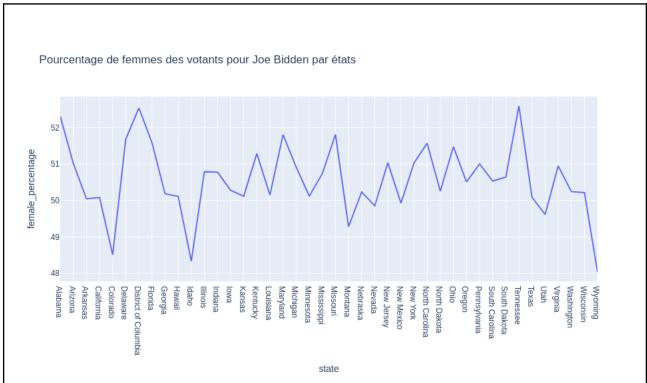
> Pourcentage de femmes dans chaque état



Ce graphique de nuage de points comporte la distribution du pourcentage des femmes dans chaque état, d'après le graphe on voit que l'intervalle du pourcentage est situé entre (48% et 53%). En comptant le nombre d'états où le pourcentage > 50%, on trouve 19/43 états et le nombre d'états où le pourcentage < 50%, on trouve 24/43 états.

On conclure que les états unis d'amérique, en particulier, dans ces 43 états, la population masculine est plus importante que la population féminine.

#### > Pourcentage de femmes des votants pour Biden dans chaque état.

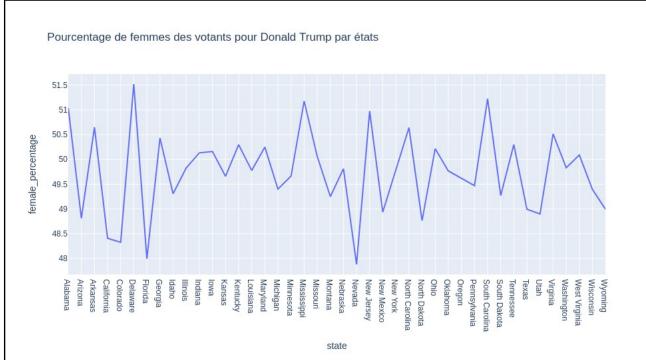


Ce graphique linéaire définit le pourcentage de femmes de la population votante pour Joe Biden,

On voit nettement que:

- L'intervalle du pourcentage de femmes des votants pour Joe Biden est situé entre 48% et 52%.
- Le pourcentage de femmes des votants pour Joe Biden est superieur à 50% dans la plupart des états (35/43 des états).

#### > Pourcentage de femmes des votants pour Trump dans chaque état.



Ce graphique linéaire indique le pourcentage de femmes de la population votante pour Joe Biden,

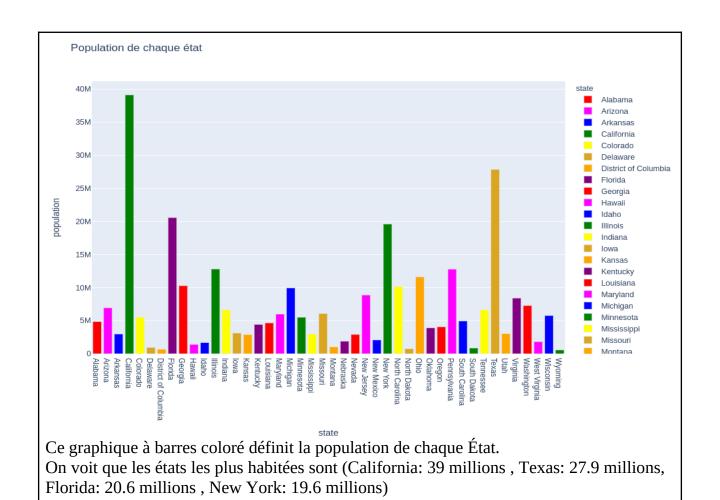
On voit d'une manière distincte que:

- L'intervalle du pourcentage de femmes des votants pour Donald Trump est situé entre 47.7% et 51.5%
- Le pourcentage de femmes des votants pour Donald Trump est inférieur à 50% dans la plupart des états (25/43 états).

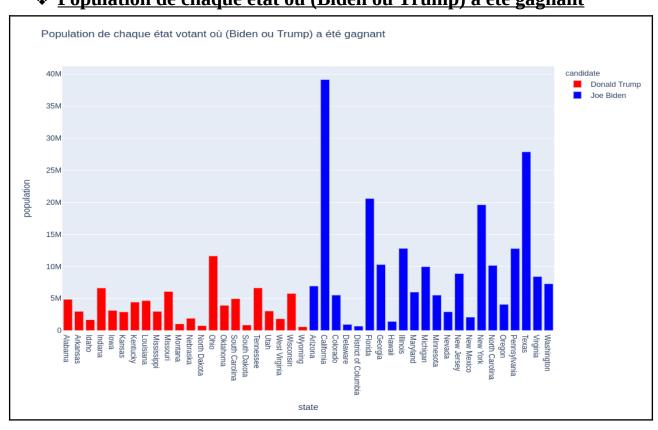
Après avoir consulté ces deux derniers graphiques linéaires, on apprend que le pourcentage de femmes dans la population votante pour Joe Biden est plus important que celle de Donald Trump. Donc, plus la population féminine est élevée dans un état, plus Joe Biden est gagnant dans cet état.

# Étude des élections par la population

❖ Affichage de la population de chaque état :



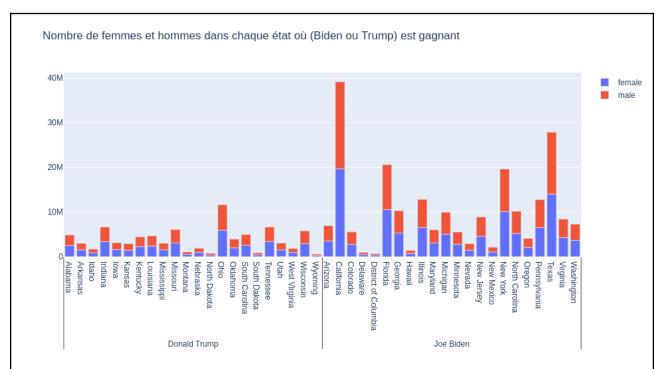
# ❖ Population de chaque état où (Biden ou Trump) a été gagnant



Ce graphique à barres comporte 43 barres qui indique la population de chaque états, ces barres sont colorées en 2 couleurs:

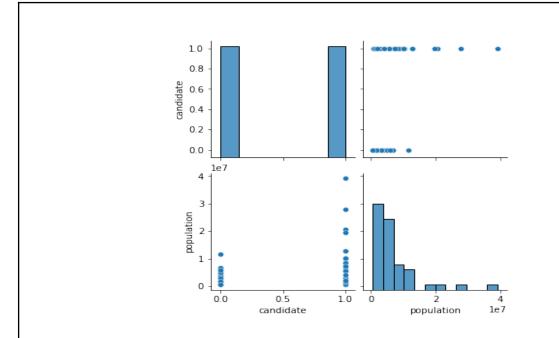
- ➤ Bleu si Joe Biden gagne dans cet état et rouge si Donald Trump est gagnant.
- ➤ On remarque que le nombre d'états gagnés est presque équivalent (Biden: 22 | Trump: 21).
- ➤ On remarque que dans la plupart du temps, Joe Biden est gagnant dans les États où la population est plus élevée comme (California, Florida, Illinois, New York, Pennsylvania, Texas). Contrairement à Trump qui a gagné dans les villes qui comportent une population très basse.

# **❖** Population par sex (masculine et féminine) dans chaque état où (Biden ou Trump) est gagnant :



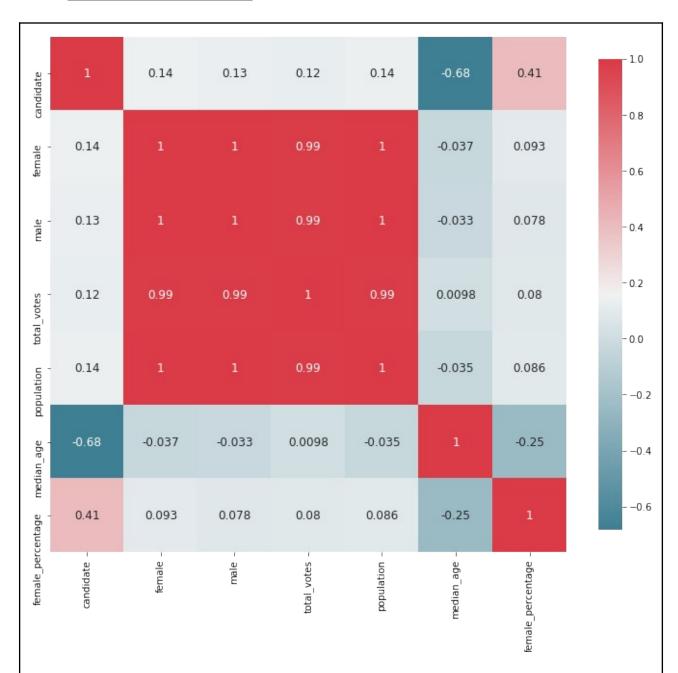
- > Ce graphique est comme celui d'avant, par contre il contient une autre information, il indique la population de chaque état, par candidat gagnant et par sex.
- > On voit que souvent, le nombre de femmes et supérieur à celui des hommes dans tous les états gagnés par les deux candidats.

# **Corrélation entre Population et candidat :**



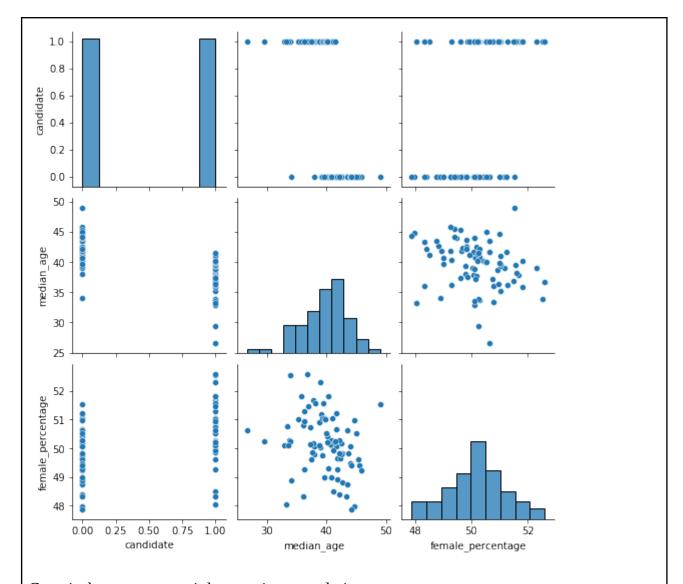
- > Ce pair plot explique une petite corrélation entre la population et le candidat,
- ➤ Ici par exemple on voit que quand candidate vaut (1 et 0) la population n'est pas élevée, par contre quand candidate vaut 1 (Joe Biden), la population est très élevée.
- > Cela montre qu'il existe une certaine corrélation entre la population et le candidat, Plus le candidat est Joe Biden, plus la population est élevée.

# **A Carte de corrélation :**



Cette map représente toutes les corrélations existantes dans la base de données. On voit pas mal de corrélations comme (male - female), (population - male), (population - female), (candidate - median\_age), (candidate - pecentage\_female)...

# **Corrélation entre candidat, âge moyen, female pourcentage :**



Ce pairplot montre aussi deux petites correlations

#### > Entre "candidate" et "median\_age":

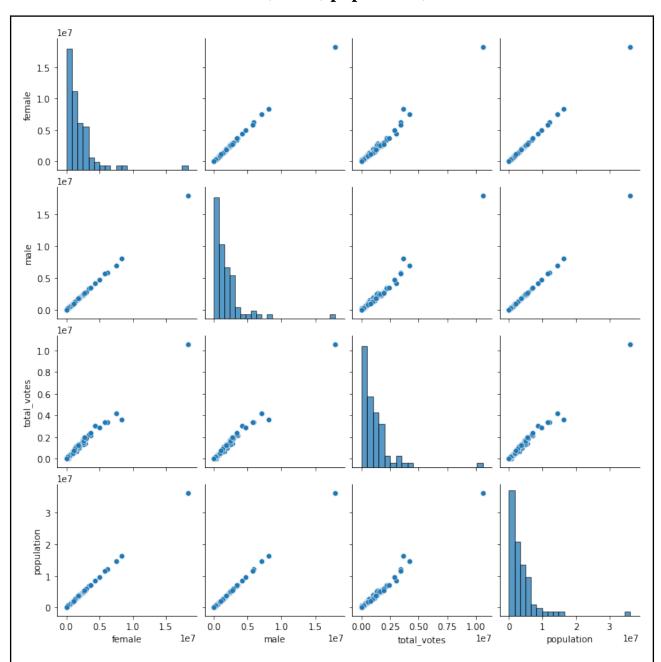
On remarque que plus le "median\_age" diminue, plus "candidate" vaut 1 qui signifie (Joe Biden) et plus le "median\_age" devient élevé plus "candidate" vaut 0 qui signifie aussi (Donald Trump)

#### > Entre "candidate" et "female\_percentage":

On voit clairement que dans le cas où "candidate" vaut 0 (Trump), "female\_percentage"

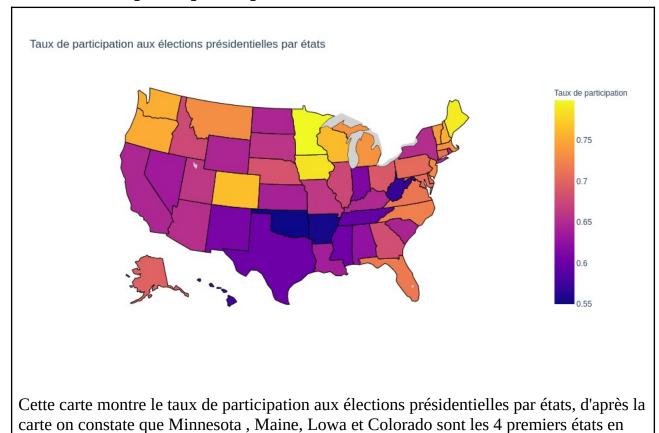
est située entre 48% et 51.5% et plus candidate vaut 1 (Joe Biden), plus "female\_percentage" augmente jusqu'à 53%.

# **Corrélation entre female, male, population, total vote :**



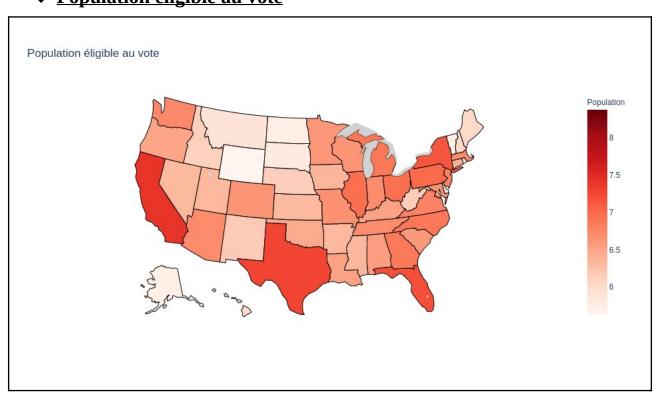
- > Ce pairplot montre qu'il y a une corrélation très forte entre total votes et les 3 autres valeurs (population, male, female).
- > On constate que plus le nombre de male, female et population augmente, plus le nombre total\_votes est élevé.

# **Taux de participation par états**



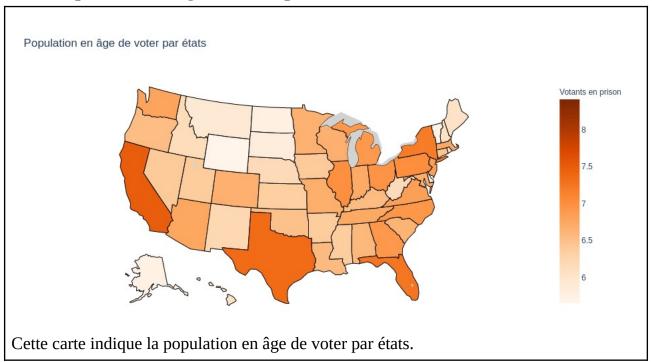
# **❖** Population éligible au vote

termes de taux de participation.



En terme de population éligible au vote, on voit que la Californie est en tête avec 25962648 personnes suivie par Texas avec 18784280 ainsi que Florida en 3 eme position avec une population de 15,5 millions personnes.

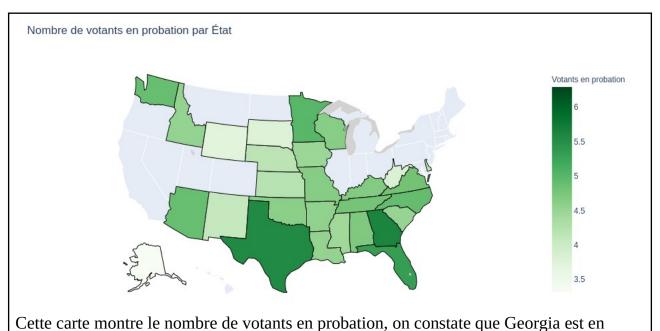
# **❖** Population en âge de voter par états



### Nombre de votants en prison par états

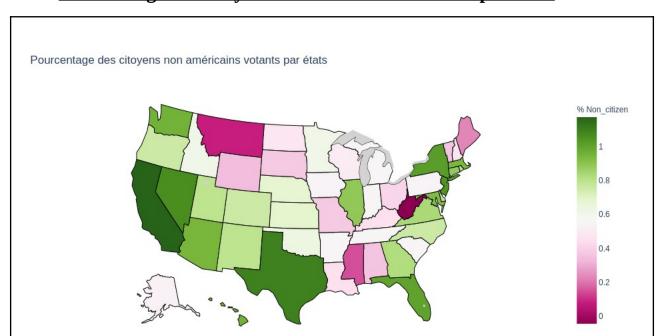


# Nombre de votants en probation par État



#### **Pourcentage des citoyens non américains votants par états**

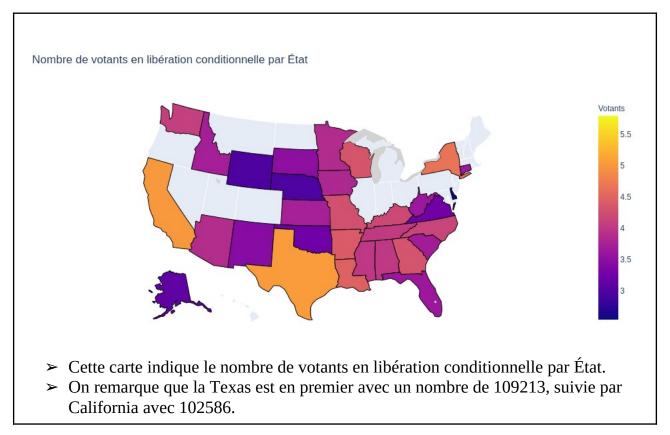
premier avec un nombre de 416771, suivie de la Texas en deuxième place avec 368167.



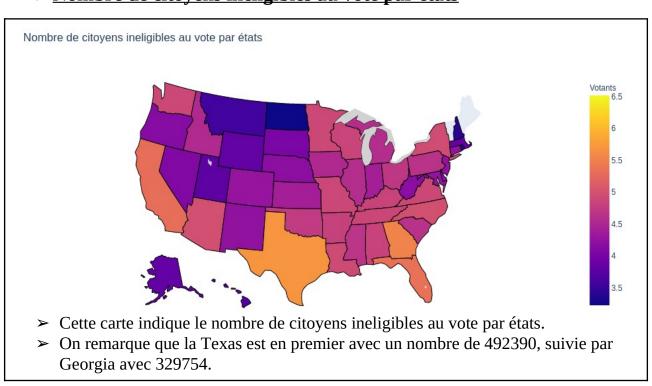
Cette carte définit le pourcentage de votants qui ne sont pas des citoyens américains, on voit clairement que les états où les grandes villes se situent sont dans les premières places en commençant par la Californie avec 15%, Texas: 12.6, Nevada: 11.6, New Jersey: 11.1 ...)

- On comprend que les grandes villes américaines portent un nombre important de citoyens non américains.

# Nombre de votants en libération conditionnelle par État



# **❖** Nombre de citoyens ineligibles au vote par états



# 4. Conclusion:

- Après avoir fait cette étude sur les élections présidentielles américaines 2020, on constate que malgré que Donald Trump a gagné un nombre de départements très supérieur à celui de Joe Biden, ce dernier a remporté les élections présidentielles.
- La population votante pour Joe Biden est plus jeune que celle de Trump.
- La population votante pour Biden est plus féminine que masculine. En revanche, celle de Trump est plus masculine.
- Joe Biden a remporté les élections grâce aux villes et états ayant un pourcentage plus élevé de la population et de citoyens non-américains, plus particulièrement California, New York, Florida et Texas.
- Grâce à ce travail, j'ai appris à manipuler beaucoup d'outils de visualisation de données en plus de ce qu'on a appris pendant toutes les 10 séances du cours.
- J'ai choisi de travailler en Python car il est très demandé dans le domaine de la science de données et fouille de données (les domaines qui m'intéressent). Par conséquent, ça me permet d'avancer et d'approfondir mes connaissances dans ce langage.