

S03 T01: Visualització gràfica d'un dataset

Descripció: Complementa les tècniques d'exploració de les dades mitjançant la visualització gràfica, amb les llibreries Matplotlib i Seaborn.

Nivell 1

Exercici 1

Resumeix gràficament el data set DelayedFlights.csv

Crea almenys una visualització per:

- Una variable categòrica (UniqueCarrier)
- Una variable numèrica (ArrDelay)
- Una variable numèrica i una categòrica (ArrDelay i UniqueCarrier)
- Dues variables numèriques (ArrDelay i DepDelay)
- Tres variables (ArrDelay, DepDelay i UniqueCarrier)
- Més de tres variables (ArrDelay, DepDelay, AirTime i UniqueCarrier).

In [12]:

```
# Crido a les llibreries de pandas, numpy, seaborn i matplotlib per poder treballar tran
# Faig entrar l'arxiu CSV gràcies a pandas

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

delayedFlightsAmbNaN = pd.read_csv(r'C:\Users\Anna\DataScience\SPRINTS\SPRINT 3\Sprint 3 T

# Elimino els NaN per fer el dataset algo més petit
delayedFlights = delayedFlightsAmbNaN.dropna()

print(delayedFlights)
```

	Unnamed: 0	Year	Month	DayofMonth	DayOfWeek	DepTime	CRSDepTime	\
3	4	2008	1	3	4	1829.0	1755	
5	6	2008	1	3	4	1937.0	1830	
7	11	2008	1	3	4	1644.0	1510	
9	16	2008	1	3	4	1452.0	1425	
11	18	2008	1	3	4	1323.0	1255	
...	
1936751	7009705	2008	12	13	6	921.0	830	
1936752	7009709	2008	12	13	6	1552.0	1520	
1936753	7009710	2008	12	13	6	1250.0	1220	
1936754	7009717	2008	12	13	6	657.0	600	
1936755	7009718	2008	12	13	6	1007.0	847	

	ArrTime	CRSArrTime	UniqueCarrier	...	TaxiIn	TaxiOut	Cancelled	\
3	1959.0	1925	WN	...	3.0	10.0	0	
5	2037.0	1940	WN	...	3.0	7.0	0	
7	1845.0	1725	WN	...	6.0	8.0	0	
9	1640.0	1625	WN	...	7.0	8.0	0	
11	1526.0	1510	WN	...	4.0	9.0	0	
...	
1936751	1112.0	1008	DL	...	8.0	21.0	0	

1936752	1735.0	1718	DL	...	9.0	7.0	0
1936753	1617.0	1552	DL	...	9.0	18.0	0
1936754	904.0	749	DL	...	15.0	34.0	0
1936755	1149.0	1010	DL	...	8.0	32.0	0

	CancellationCode	Diverted	CarrierDelay	WeatherDelay	NASDelay	\
3	N	0	2.0	0.0	0.0	
5	N	0	10.0	0.0	0.0	
7	N	0	8.0	0.0	0.0	
9	N	0	3.0	0.0	0.0	
11	N	0	0.0	0.0	0.0	
...	
1936751	N	0	51.0	0.0	13.0	
1936752	N	0	0.0	0.0	0.0	
1936753	N	0	3.0	0.0	0.0	
1936754	N	0	0.0	57.0	18.0	
1936755	N	0	1.0	0.0	19.0	

	SecurityDelay	LateAircraftDelay
3	0.0	32.0
5	0.0	47.0
7	0.0	72.0
9	0.0	12.0
11	0.0	16.0
...
1936751	0.0	0.0
1936752	0.0	17.0
1936753	0.0	22.0
1936754	0.0	0.0
1936755	0.0	79.0

[1247486 rows x 30 columns]

- Una variable categòrica (UniqueCarrier)

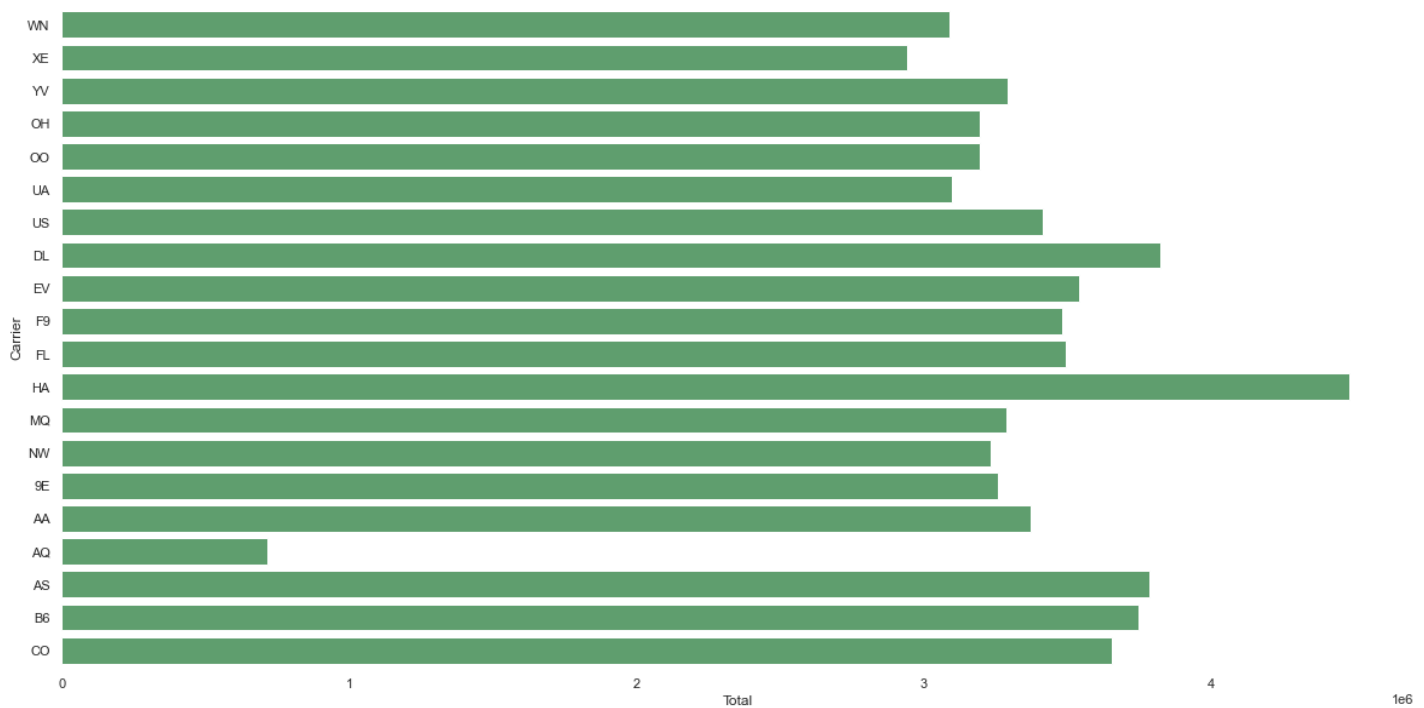
In [13]:

```
# Inicialitzo la figura que vull
f, unCarr = plt.subplots(figsize=(20, 10))

# Canvio el color de fons
sns.set(style="white")

# Creo el gràfic, en aquest cas un barplot
img1 = sns.barplot(
    x="Unnamed: 0",
    y="UniqueCarrier",
    data=delayedFlights,
    label= "",
    ci=None,
    orient= "h",
    color="g");

# Afegeixo una llegenda
unCarr.set(
    ylabel="Carrier",
    xlabel="Total")
sns.despine(left=True, bottom=True)
```



- Una variable numèrica (ArrDelay)

In [14]:

```
# Inicialitzo la figura que vull
x, ad = plt.subplots(figsize=(40, 10))

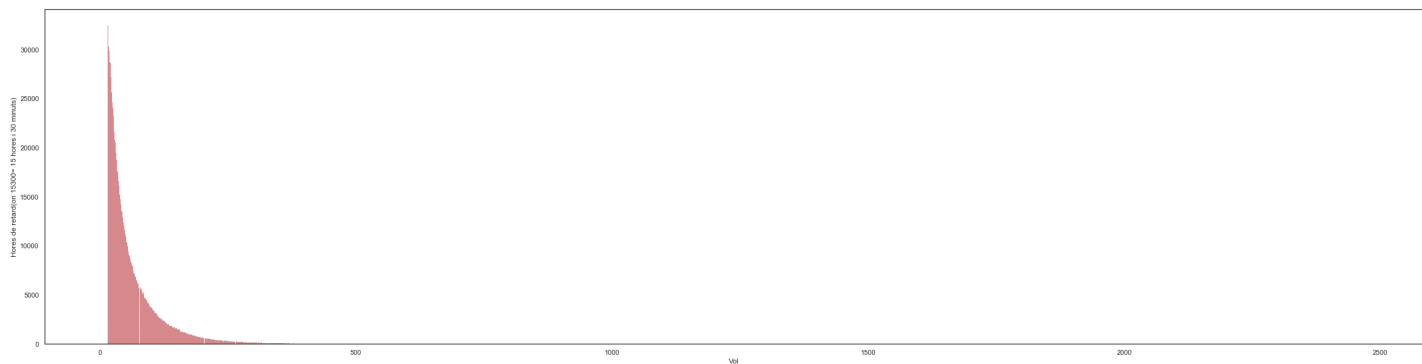
# Canvio el color
sns.set(style="dark")

# Creo el gràfic, en aquest cas un histograma
img2= sns.histplot(
    #bins=100,
    x="ArrDelay",
    data=delayedFlights,
    color="r");

# Afegeixo una llegenda
ad.set(
    ylabel="Hores de retard(on 15300= 15 hores i 30 minuts)",
    xlabel="Vol")
```

Out[14]:

```
[Text(0, 0.5, 'Hores de retard(on 15300= 15 hores i 30 minuts)'),
 Text(0.5, 0, 'Vol')]
```



- Una variable numèrica i una categòrica (ArrDelay i UniqueCarrier)

In [15]:

```
# Inicialitzo la figura que vull
x, adUC = plt.subplots(figsize=(20, 30))
```

```
# Canvio el color
sns.set(style="white")

# Creo el gràfic, en aquest cas un de violins
img3 = sns.violinplot(
    y="UniqueCarrier",
    x="ArrDelay",
    data=delayedFlights,
    cut=0);

# Afegeixo una llegenda
adUC.set(
    ylabel="Unique Carrier",
    xlabel="Retard d'arribada")
```

Out[15]: [Text(0, 0.5, 'Unique Carrier'), Text(0.5, 0, "Retard d'arribada")]



- Dues variables numèriques (ArrDelay i DepDelay)

In [16]:

```
# Inicialitzo la figura que vull
x, adDD = plt.subplots(figsize=(15, 15))

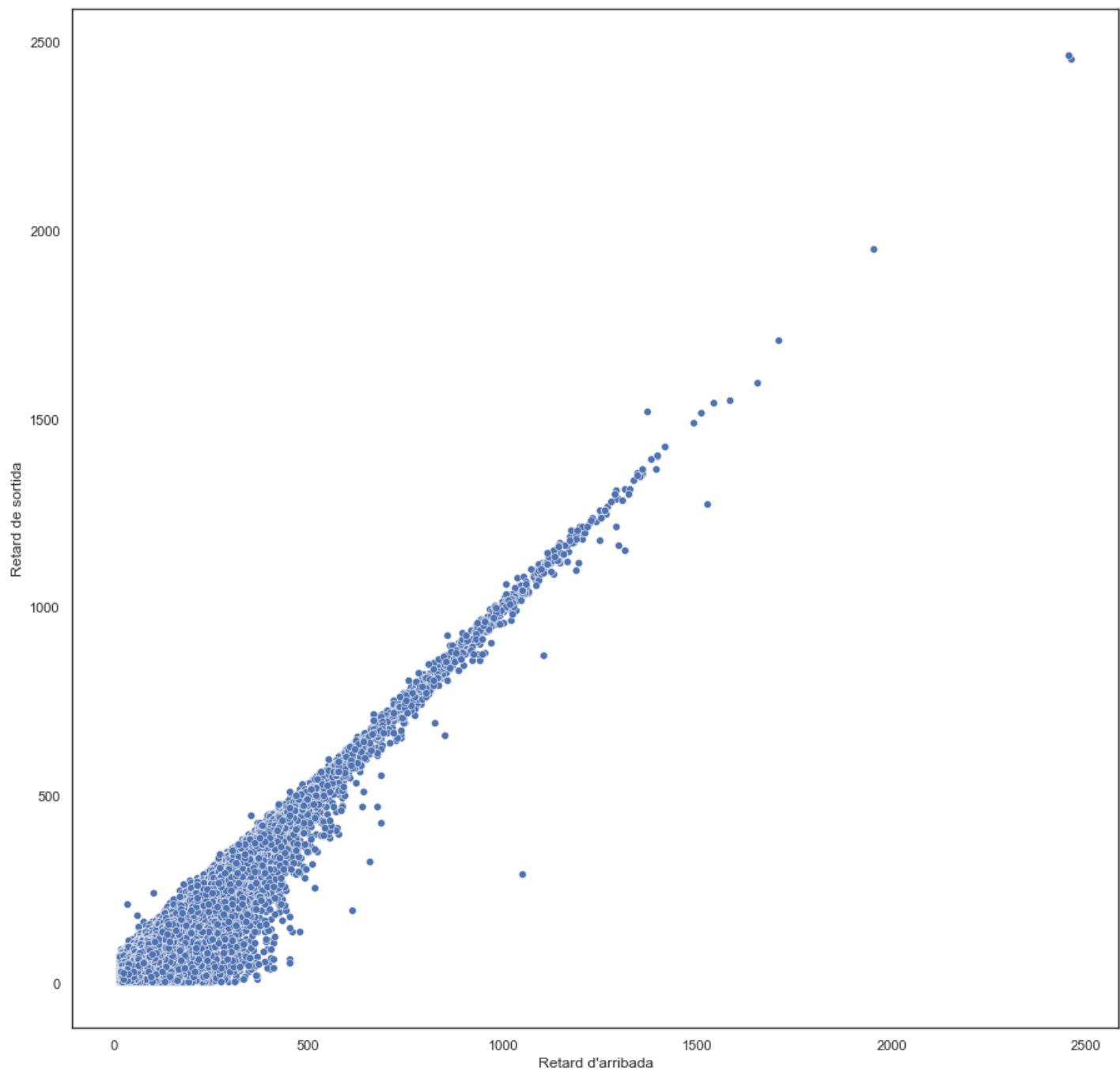
# Canvio el color
sns.set(style="white")

# Creo el gràfic, en aquest cas un scatterplot
img4 = sns.scatterplot(
    y="DepDelay",
    x="ArrDelay",
    data=delayedFlights,
    color="b",
    marker='o',
    markers=0.5);

# Afegeixo una llegenda
adDD.set(
    ylabel="Retard de sortida",
    xlabel="Retard d'arribada")
```

Out[16]:

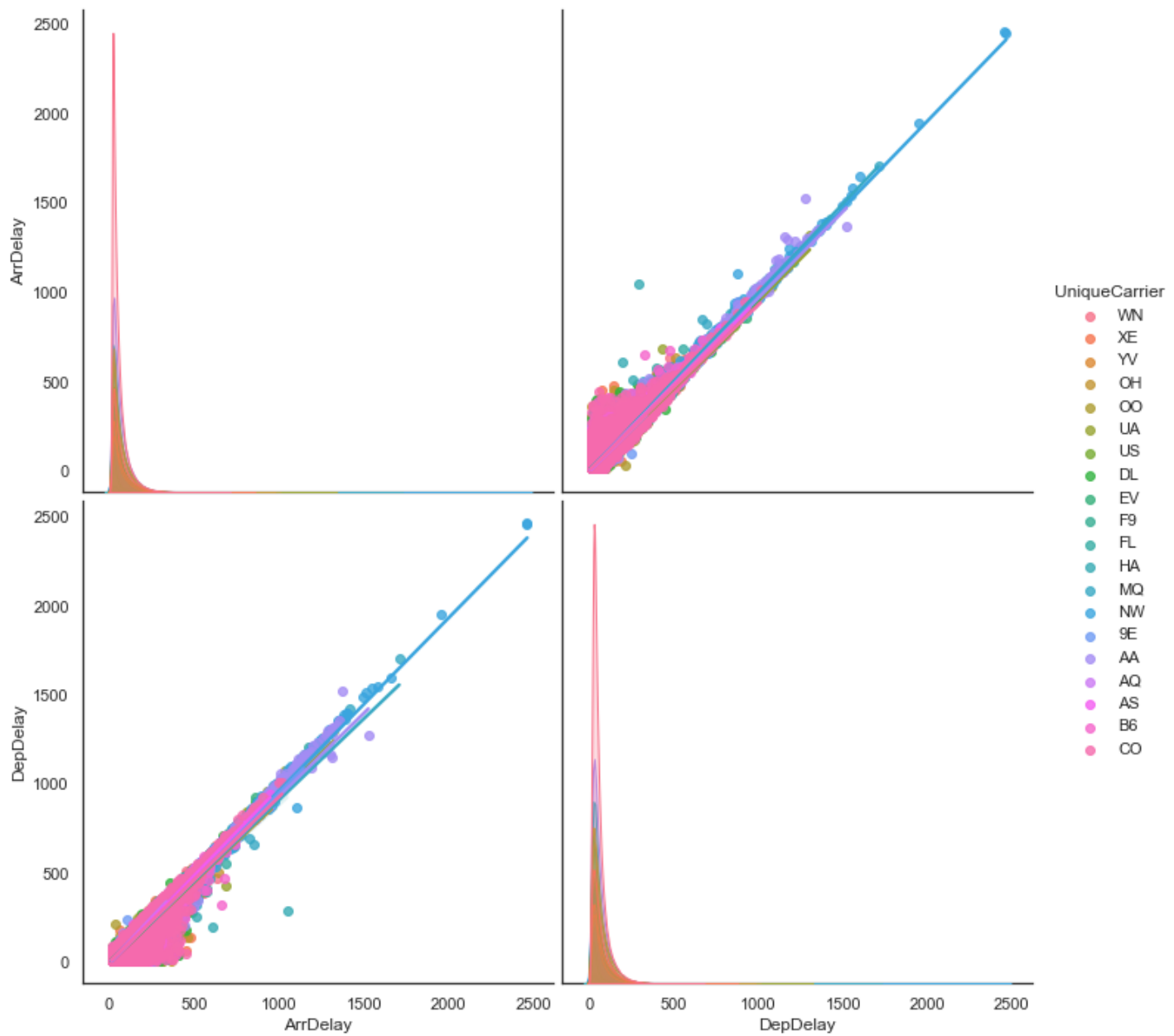
```
[Text(0, 0.5, 'Retard de sortida'), Text(0.5, 0, "Retard d'arribada")]
```



- Tres variables (ArrDelay, DepDelay i UniqueCarrier)

In [17]:

```
# Creo el gràfic, en aquest cas un correlogram (pairplot)
img5 = sns.pairplot(
    data=delayedFlights,
    x_vars=["ArrDelay", "DepDelay"],
    y_vars=["ArrDelay", "DepDelay"],
    hue= "UniqueCarrier",
    kind="reg",
    height=5);
```



```
In [ ]: fig5a = sns.stripplot(
    x='ArrDelay',
    y='DepDelay',
    data= delayedFlights,
    hue= "UniqueCarrier",
    color="orange",
    jitter=0.2,
    size=2.5);
```

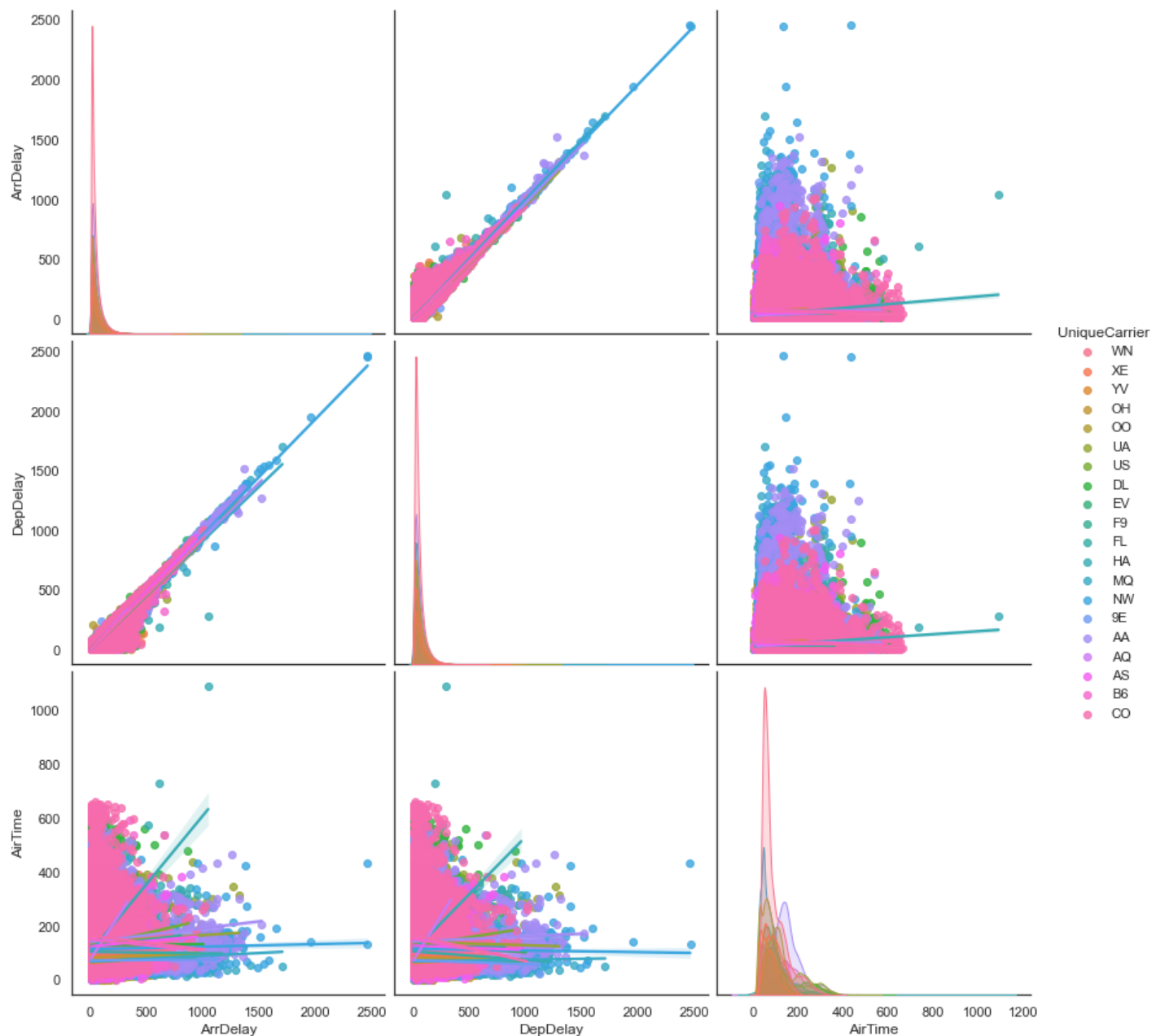
```
In [18]: '''
sns.jointplot(
    data=delayedFlights,
    x="ArrDelay",
    y="DepDelay",
    hue="UniqueCarrier",
    kind="kde");
'''
```

```
Out[18]: '\nsns.jointplot(\n    data=delayedFlights,\n    x="ArrDelay", \n    y="DepDelay", \n    hue="UniqueCarrier",\n    kind="kde");\n'
```

- Més de tres variables (ArrDelay, DepDelay, AirTime i UniqueCarrier).

In [19]:

```
# Creo el gràfic, en aquest cas un correlogram (pairplot)
img6 = sns.pairplot(
    data=delayedFlights,
    x_vars=["ArrDelay", "DepDelay", "AirTime"],
    y_vars=["ArrDelay", "DepDelay", "AirTime"],
    hue= "UniqueCarrier",
    kind="reg",
    height=4);
```



Exercici 2

Exporta els gràfics com imatges o com html.

In [25]:

```
fig = img1.get_figure()
fig.savefig("img1.png")

fig = img2.get_figure()
fig.savefig("img2.png")

fig = img3.get_figure()
fig.savefig("img3.png")
```

```
fig = img4.get_figure()
fig.savefig("img4.png")

# Les imatges següents no m'anava el get_figure i se m'han guardat així
img5.savefig("img5.png")
img6.savefig("img6.png")
```

Exercici 3

Integra les visualitzacions gràfiques, en la tasca 5, del Sprint 2.

Mirar Spring 2 T05, al final

In []: