



Hópaverkefni 1

Gagnavinnsla

Davíð Halldórsson
Fannar Örn Arnarsson
Jóhannes Hilmarsson
Ómar Sindri Jóhannsson

Dagsetning skýrslugerðar: 4. desember 2015
Kennari: Eyjólfur Ingi Ásgeirsson

1 Rannsóknarspurning

Er einhver tilviljunarkenndur þáttur sem hefur áhrif á afbrotatíðni á Íslandi? Getum við sýnt fram á það með gögnum frá Hagstofunni að eitthvað samband sé á milli þessara þátta sem við tökum fyrir í þessari rannsókn og afbrotatíðni?

2 Tilgáta

Teknir verða fyrir nokkrir þættir og borið þá saman við fjölda afbrota á Íslandi. Við ákváðum að taka tilviljunakennd gögn frá Hagstofunni og finna mestu fylgnina á milli afbrota og gagnanna og setja fram kenningu um hvort gögnin hafi raunveruleg áhrif á fjölda afbrota.

3 Forsendur

Við gefum okkur það að fylgni geti sýnt raunverulega mynd af áhrifum hvers þáttar á fjölda afbrota. Kóðinn okkar gerir ráð fyrir því að .csv skrárnar innihalda dálk sem heitir „Ár“, þær hafi „Encoding“utf-8 og hvert ár komi aðeins fram einu sinni ásamt því að skjalið sé án töfluheitis.

4 Aðferð

Forritið er hannað til að geta lesið inn nánast hvaða tvær .csv skrár af Hagstofunni og borið þær saman. Þar á eftir tekur forritið þessi tvö skjöl og finnur sameiginlegt tímabil og setur í eitt „dataframe“.

Listing 1: Kóði úr forriti.

```
def Lesaskra(nafn):
    settings = {'encoding': 'utf-8', 'sep': ';', 'dayfirst': True}
    skjal = pd.read_csv(nafn, **settings)
    skjal.set_index('Ar', inplace=True)
    skjal = skjal.apply(pd.to_numeric, errors='coerce').dropna(axis=1, how='all')
    return skjal

Valinskra=input('Veldu skra til að bera saman við "Afbrot.csv": ')
skjal1=Lesaskra('Afbrot.csv')
skjal2=Lesaskra(Valinskra)
skjal = pd.concat({'mannfjöldi.csv': skjal1, 'Afbrot.csv': skjal2}, axis=1, join_axes=[skjal1.index])
```

Út frá gögnunum reiknar forritið út grunntölfræðigögn eins og meðaltal og stalaðfrávik þó að litið sé burt frá þeim við ályktun á niðurstöðum. Notandi velur dálka úr .csv skránum og forritið finnur prósentubreytingu milli ára og notar þau gögn til að teikna upp graf og reikna fylgnina.

Listing 2: Kóði úr forriti (aðlagður að skýrsluformi).

```
d = {}
d['Stadalfravik'] = np.std(skjal, axis=0)
d['Medaltal'] = np.mean(skjal, axis=0)
d['Midgildi'] = np.median(skjal, axis=0)
d['Max'] = np.max(skjal, axis=0)
d['Min'] = np.min(skjal, axis=0)
df = pd.DataFrame(d)

[print(i, end=";\n") for i in skjal]
x1 = input("\nVeldu skra:\n")
x2 = input("\nVeldu undirflokk:\n")
y1 = input("\nVeldu skra:\n")
y2 = input("\nVeldu samanburðarflokk:\n")

plt.style.use('ggplot')
df2=pd.DataFrame({x2: skjal[x1][x2].pct_change(), y2: skjal[y1][y2].pct_change()}).dropna(axis=0)
print('Fylgni valinna gagna er:', np.corrcoef(df2[x2], df2[y2])[0][1])
df2.plot()
plt.show()

skjal.to_csv('Gogn.csv')
df.to_csv('Tolfraedi.csv')
df2.to_csv('breytingFlokka.csv')
```

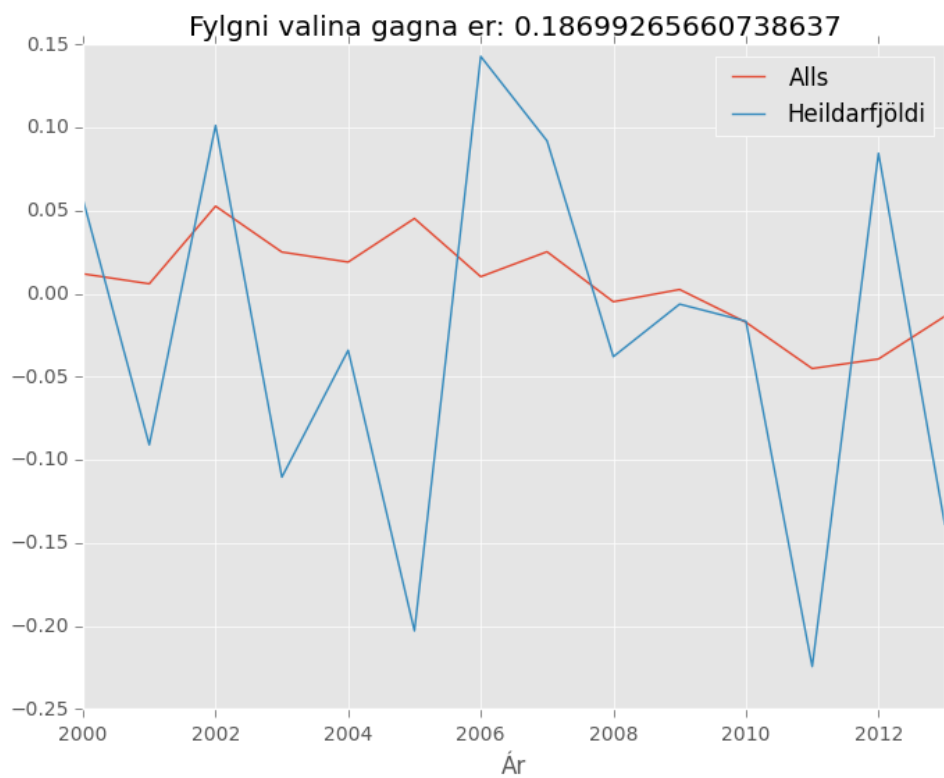
Í þessu verkefni tökum við nokkur mismunandi .csv skjöl frá Hagstofunni. Hvaða gögn voru notuð má sjá í kaflanum „Niðurstöður“ hér að neðan. Allar .csv skrárnar voru bornar saman við tölur frá Hagstofunni um fjölda afbrota á Íslandi.

5 Niðurstöður

Niðurstöður voru fengnar úr fimm keyrslum og fundið var út hvort einhver fylgni væri á milli gagnanna. Keyrslurnar voru eftirfarandi:

5.1 Keyrsla 1

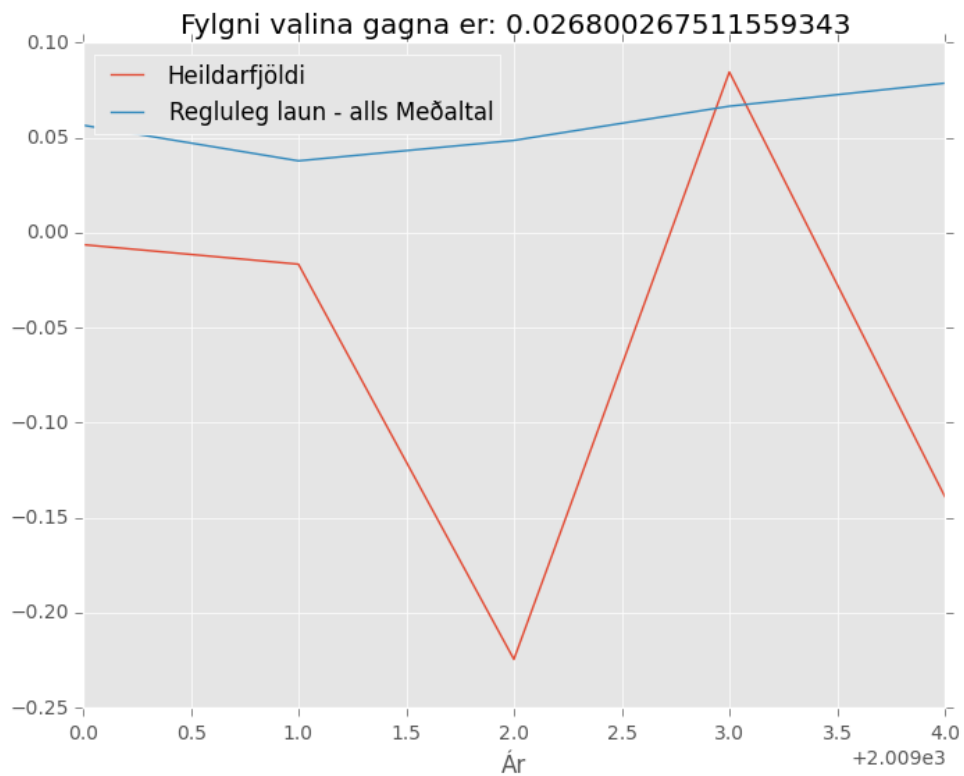
Í fyrstu keyrslunni var borið saman fjöldi afbrota og fjölda flugvéla á Íslandi.



Mynd 1: Breyting á flugvélafjölda og afbrota milli ára.

5.2 Keyrsla 2

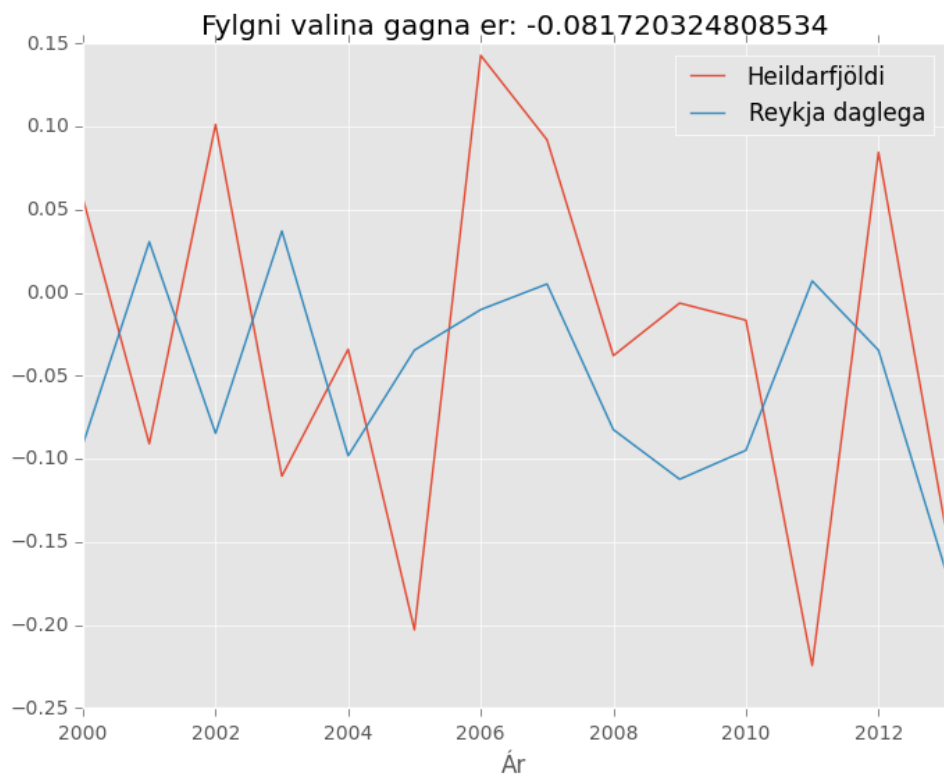
Í annarri keyrslunni var borið saman afbrot og meðallaun Íslendinga.



Mynd 2: Breyting á meðallaunum og afbrota Íslendinga milli ára.

5.3 Keyrsla 3

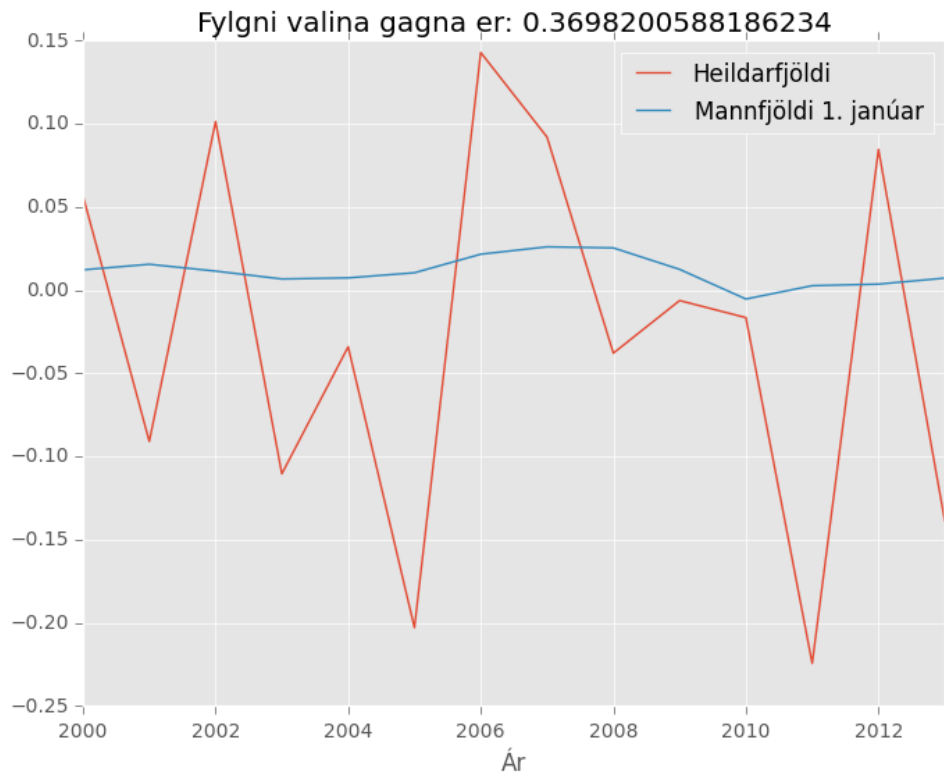
Í þriðju keyrslunni var borið saman afbrot og fjöldi dagreykingamanna á Íslandi á aldrinum 15-79 ára.



Mynd 3: Breyting á fjölda dagreykingafólks og afbrota á Íslandi milli ára.

5.4 Keyrsla 4

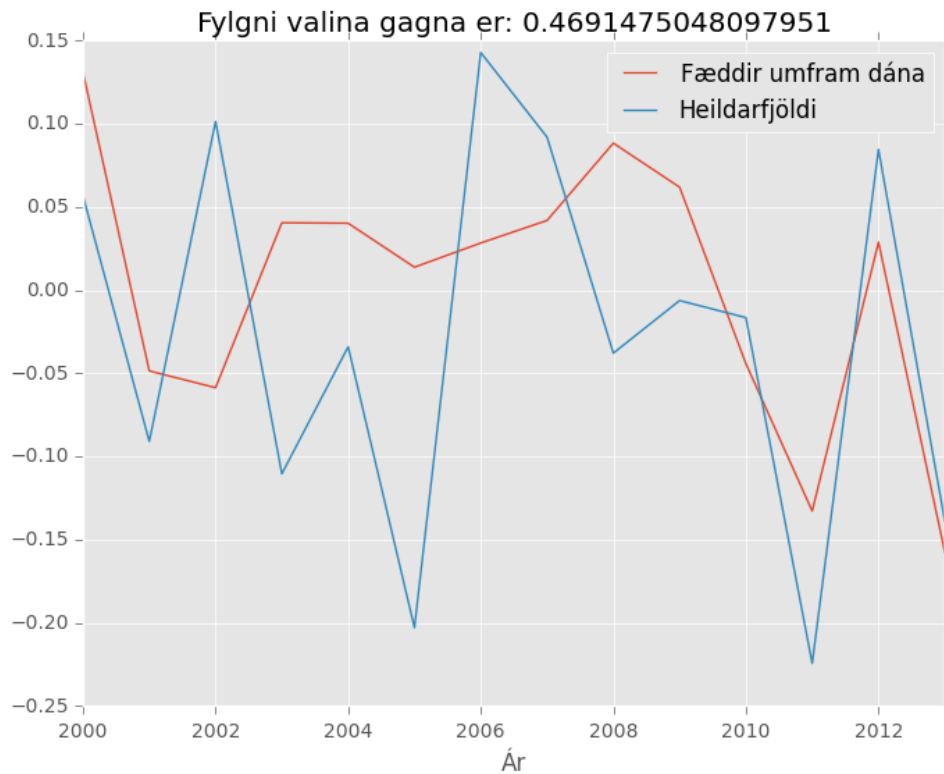
Í fjórðu keyrslunni var borið saman afbrot og mannfjölda á Íslandi.



Mynd 4: Breyting á mannfjölda og afbrota á Íslandi milli ára.

5.5 Keyrsla 5

Í síðustu keyrslunni var borið saman afbrot og fæddir umfram dána á Íslandi.



Mynd 5: Breyting á fjölda fæddra umfram dána og afbrota Íslendinga milli ára.

Eins og vænta mátti þá sést á niðurstöðunum hér að ofan að allir þessir þættir hafa lítið sem engin áhrif á fjölda afbrota á Íslandi nema fæddir umfram dána. Úr niðurstöðum úr keyrsu fimm sést að fylgnin er mun meiri en í hinum niðurstöðunum, þó ekki nógu miklar að hægt sé að draga ályktun um að þessi þáttur hafi einhver raunveruleg áhrif á fjölda afbrota á Íslandi.