



HÁSKÓLI ÍSLANDS

Gagnafrásögn - skýrsla

Targaryen: Ásdís G, Ásdís Halla, Gígja Marín og Þráinn Ágúst

Lýsing á gagnaöflun

Gagnaöflun verkefnisins byggðist á því að sækja gögn frá vefsíðunni timataka.net, sem heldur utan um úrslit úr íslenskum hlaupum. Í þessu verkefni er aðeins unnið með niðurstöður úrslita í ágúst mánuði frá árunum 2018-2024. Til að tryggja áreiðanleika gagnana voru gögnin sótt beint frá opinberum úrslitum (e. Heildarúrslit/Overall) og unnin í gegnum forritunarlausnir sem tryggja samræmi og nákvæmni í öflun og vistun gagna.

Gögnin voru sótt með Python forritum sem byggja á verkefninu frá Martell hópnum, úr SQLite verkefninu í viku 6. Notuð voru þrjú lykilforrit: `timataka.py`, `agust_url.py`, og SQL skráin `create_db.sql` ásamt textaskránni `urls.txt`. Hér er frekari lýsing á ferlinu:

- **Timataka.py:** Þetta forrit var notað til að sækja gögn um úrslit keppenda beint af `timataka.net`. Forritið les HTML kóða af vefslóðunum og vinnur úr þeim lykilupplýsingar, eins og nafn hlaupsins, fjölda keppenda, upphafstíma, og úrslit keppenda. Gögnin eru síðan vistuð í tveimur CSV skráum, þar sem önnur heldur utan um upplýsingar um hlaupin (`hlaup_info.csv`) og hin geymir úrslit keppenda (`hlaup.csv`). Forritið tryggir nákvæma öflun gagna með því að nota reglulegar segðir til að greina gögnin úr HTML kóðanum.
- **August.py:** Þetta forrit var notað til að sækja gögnin í einu skrefi fyrir öll hlaup sem tilheyra ágúst mánuði 2018-2024. Slóðirnar fyrir þessi hlaup voru geymdar í textaskránni `urls.txt`. Forritið keyrir `timataka.py` í röð, fer í gegnum listann af slóðum, og sækir gögn um hvert hlaup.
- **Create_db.sql:** SQL skráin var notuð til að búa til SQLite gagnagrunninn `timataka.db`. Í þeim gagnagrunni verða til tvær töflur, `timataka` og `hlaup`. Taflan `timataka` heldur utan um helstu upplýsingar um keppendur og taflan `hlaup` heldur utan um upplýsingar um hlaupin. Eftir keyrslu var gagnagrunnurinn notaður til að geyma og vinna með gögnin sem voru sótt með Python forritunum.

Við frekari vinnslu gagna kom í ljós að sumir dálkar í CSV skránum höfðu villur, eins og tóma dálka. Þessi gögn voru flutt yfir í R til frekari hreinsunar. Þar voru auka dálkar fjarlægðir og endanleg gögn voru síðan vistuð aftur sem CSV skrár.

Við tengdumst gagnagrunninum í R með þökkunum DBI og RSQLite, og tókum þannig við frekari úrvinnslu og greiningu á gögnunum.

Gagnavinnsla

Hreinsun á gögnum

Þegar að gögnin voru komin yfir í R fór frekari skoðun í gang á töflunni `timataka_data`, sást þá að á sumum stöðum voru tómar raðir sem innihéldu aðeins `id` og `hlaup_id`. Þeim röðum var hreinsað út og gert var ný tafla sem heitir `timataka_data2`. Með hreinan gagnagrunn sem hægt væri að nota til frekari úrvinnslu á gögnunum hófst vinna við val á verkefni.

Val og birting á efni og gröfum

Til að finna áhugaverða einstaklinga til að fjalla um var skoðað hvaða einstaklingar voru oftast að keppa. Til þess að finna það var notað `count` til að telja hvað allir einstaklingar voru að keppa oft og síðan aðeins prentað út og geymt í sér töflu þá einstaklinga sem komu oftar en 14 sinnum fram. Aðeins átta einstaklingar voru oftar en 14 sinnum og af þeim voru þau Bergdís Eva Sveinsdóttir og Eyjólfur Guðgeirsson sem kepptu oftast eða 17 sinnum. Rúnar Sigurðsson keppti einnig 17 sinnum en ákváðum við að hafa aðeins þau Bergdís og Eyjólf eftir að skoðað var hvar þau enduðu í keppnum hjá sér, `Rank`, sem sagt að þau enduðu hærra heldur en Rúnar.

Eftir að búið var að ákveða að hafa þau Bergdís og Eyjólf, þá var farið í það að búa til sér gagnatöflur fyrir þau bæði eða `bergdis_data` og `eyjolfur_data` sem innihélt aðeins upplýsingar um þeirra keppnir úr `timataka_data2`.

Til að geta borið þau mögulega saman þá þarf að finna hvort þau hafa keppt í sömu keppnum. Eftir að hafa keyrt þann kóða kom í ljós að þau höfðu aðeins keppt þrisvar sinnum í sömu keppnum. (Sjá mynd fyrir neðan)

id	nafn
224	KIA Gullhringurinn 2019 - KIA Gullhringurinn 106 km
228	Fellahringurinn 2019 - Stóri Hringur - 29 KM
312	Castelli Classic TT 2022 - Castelli Classic TT: 22 KM - Heildar...

Næsta skref var að gera línurit til þess að bera saman tíma þeirra í þessum hlaupum en einnig Rank s.s. hvar þau enduðu bæði tvö. Einnig vildum við sjá hvernig þau sem einstaklingar voru að ganga og þarf þá skoða þau hlaup eða hjólreiðar sem þau hafa tekið þátt í nokkrum sinnum. Við yfirlestur á `bergdis_data` og `eyjolfur_data`, þeirra eigin gagnatöflum, sást að Bergdís hefði aðeins tekið þátt í Íslandsmótinu í Criterium oft en tvisvar. Eyjolfur hefur einnig keppt oft en tvisvar í Criterium sem og Fellahringnum. Með þær upplýsingar fyrir hendi var síðan gert þrjú línurit til að sýna árangur þeirra beggja í þessum keppnum milli ára með tilliti til tíma. Tvö fyrir báðar keppnir hans Eyjólfsson og eitt fyrir hana Bergdís.

Það næsta sem við fórum að skoða var hversu oft þau Bergdís og Eyjolfur voru að keppa ár hvert. Til þess var notað `count` til að einfaldlega telja þann fjölda af keppnum hjá þeim báðum og það sett upp í stöplarit. Áður en það var hægt þurfti að draga úr nafna dálknum yfir keppninar ártölin og búa til nýjan dálk. Með árin komin var síðan sameinað `bergdis_data` og `eyjolfur_data` (þeirra eigin gagnatöflur) til þess að hafa eina töflu yfir öll gögnin og úr því búið til stöplarit.

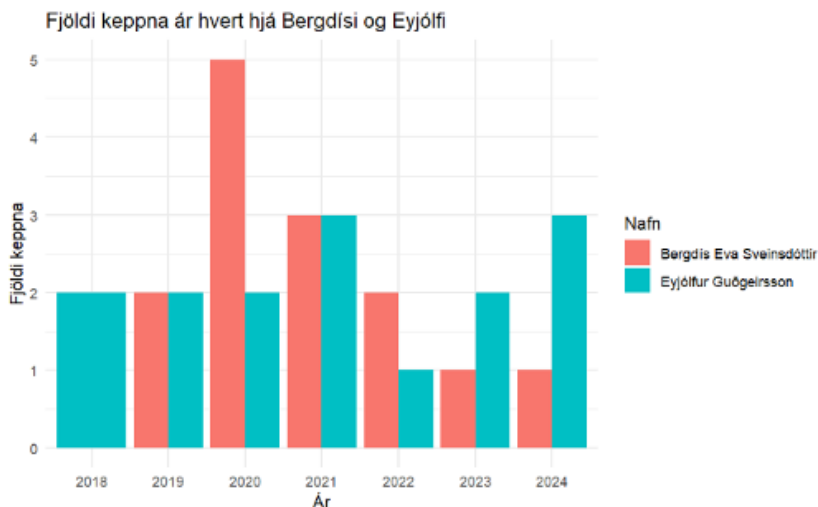
Einnig vildum við sjá hversu oft þau voru að lenda í þessum svokölluðu “top 3” sætum. Notað var talninguna yfir fjölda keppna hjá þeim og síðan aukalega talið hversu oft þau lentu í Rank númer eitt, tvö eða þrjú. Þeirri talningu var síðan deilt með fjölda keppna og út úr því kom prósentan yfir þau skipti sem þau Bergdís og Eyjolfur lentu í efstu þremur sætum. Því var síðan sett inn í stöplarit og merkt alla viðeigandi þætti.

Það síðasta sem við skoðuðum var meðaltíminn hjá þeim per kílómetur eða svokallað “pace”. Þar sem þau bæði tvö eru aðallega í hjólreiðum þá voru allar hlaupakeppnir

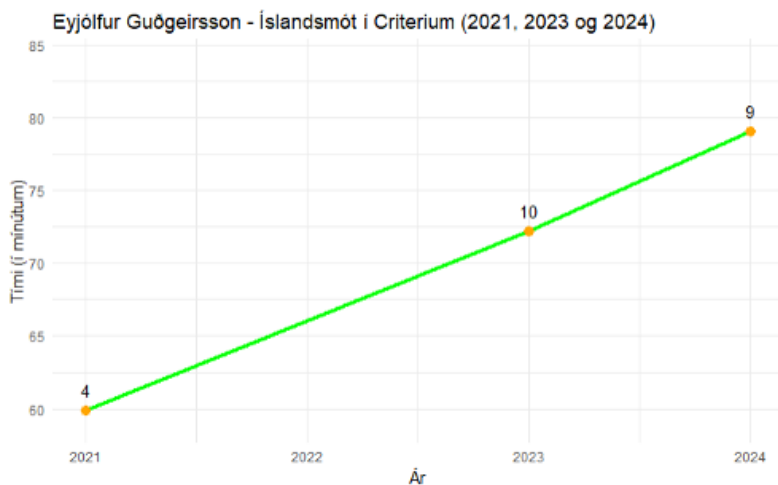
teknar út og aðeins hjólreiða keppnir notaðar. Því næst þurfti að draga úr nafna dálknum yfir heitin á keppnunum fjöldann af kílómetrum í sér dálk. Tíma uppröðunin var HH:MM:SS og því skipt yfir í aðeins mínútur. Sett var svo upp línurit sem sýndi meðaltíma per kílómeter í hverri keppni.

Niðurstöður

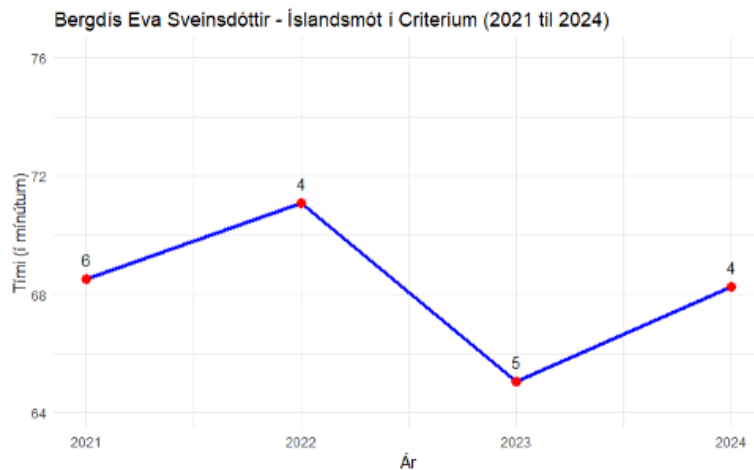
Skoðaður var fjöldi keppna sem Bergdís og Eyjólfur tóku þátt í á árunum 2018 til 2024 og sýndu niðurstöður að Bergdís hafði keppt í flestum keppnum árið 2020, eða fimm keppnum alls, og Eyjólfur hafði keppt mest í þremur keppnum á ári, árin 2021 og 2024. Bergdís keppti í fæstum keppnum árin 2023 og 2024 þar sem hún keppti bara í einni keppni á ári þau ár. Eyjólfur keppti einnig minnst í einni keppni en það var árið 2022.



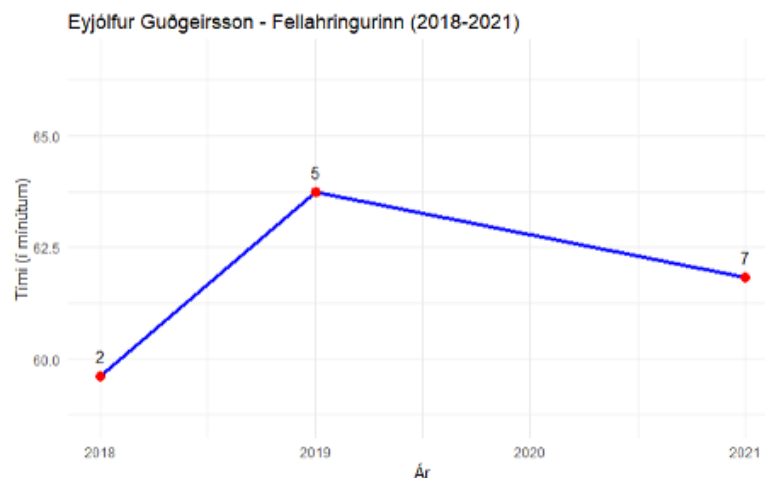
Næst var frammistaða Eyjólfur á Íslandsmóti í Criterium skoðaður á milli ára, en hann keppti í þeirri keppni árin 2021, 2023 og 2024. Besti tíminn hans var árið 2021 þegar hann kláraði á 59 mínútum og 55 sekúndum en þá lenti hann í fjórða sæti í keppninni. Tíminn hans fór hækkandi í seinni skiptin sem hann tók þátt en þá lenti hann í 10 og 9 sæti.



Frammistaða Bergdísar á sama móti var einnig skoðaður en hún keppti á því móti í fjögur skipti, 2021, 2022, 2023 og 2024. Minnsti tíminn hennar var árið 2023 en þá lenti hún í fimmta sæt. Bæði árið 2022 og 2024 lenti hún í fjórða sæti en tíminn hennar var betri árið 2024 en 2022.

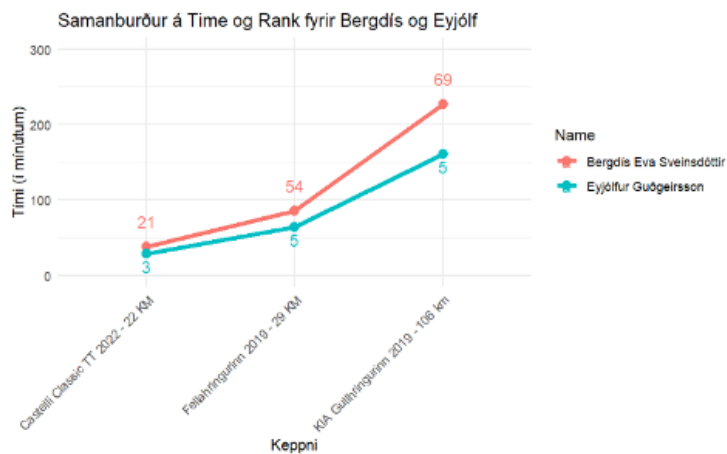


Eyjólfur tók þrisvar sinnum þátt í Fellahringnum, árin 2018, 2019 og 2021, en besti tíminn hans var árið 2018 þegar hann lenti í öðru sæti. Tíminn hans var hærri árið 2019 en þá lenti hann í fimmta sæti og svo lenti hann í sjöunda sæti árið 2021.

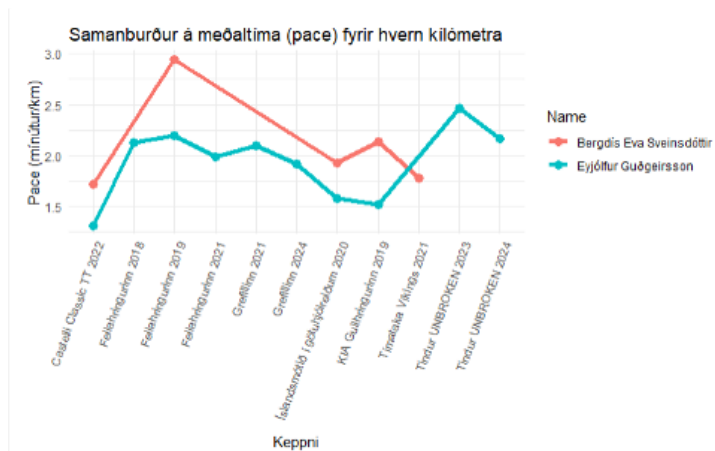


Einnig var gerður samanburður á Bergdís og Eyjólf í þremur keppnum sem þau tóku bæði þátt í en hann sýnir að Eyjólfur var yfirleitt aðeins fljótari en Bergdís. Hins vegar var munurinn

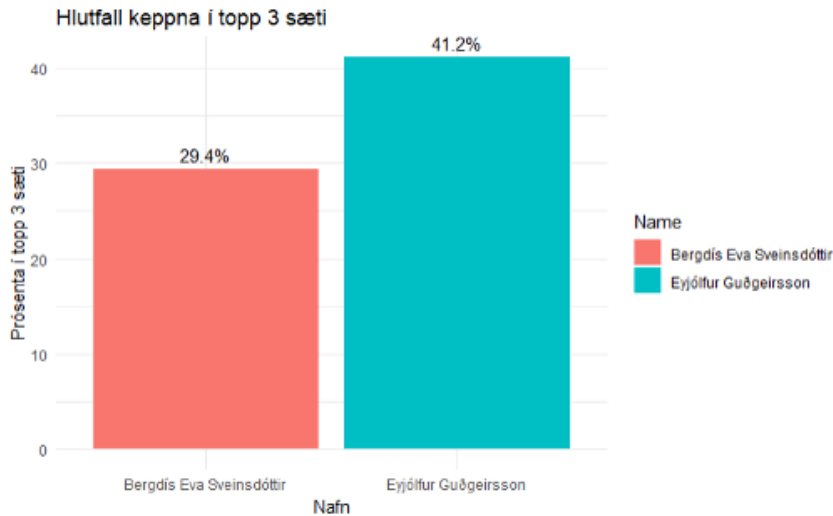
mismikill eftir keppnum. Það sést einnig að Eyjólfur lenti alltaf í mun betri sætum í keppnunum en Bergdís þar sem hann lenti í þriðja sæti og svo tvisvar í fimmta sæti á meðan hún lenti í



Myndin sýnir samanburð á hraða Bergdísar og Eyjólfss fyrir hvern kílómetra í hjóltreíðakeppnum sem þau tóku þátt í. Eyjólfur virðist hafa haldið stöðugri hraða á meðan Bergdís sýnir meiri sveiflur eftir því sem leið á keppnirnar. Þetta gæti bent til mismunandi keppnisaðferða, þar sem Eyjólfur gæti verið stöðugri í sinni nálgun á meðan Bergdís beitir meiri krafti í ákveðnum köflum keppnanna.



Að lokum var hlutfall þeirra keppna þar sem Bergdís og Eyjólfur lentu í einu af efstu þremur sætunum skoðað og borið saman. Bergdís lenti í efstu þremur sætunum í 29.4% keppna sem hún tók þátt í en Eyjólfur í 41.2%.



Mælaborðið

Tilgangur mælaborðsins er að veita notendum innsýn í frammistöðu yfir tíma og bera saman árangur tveggja hlaupara og hjólreiðafólks, Bergdísar og Eyjólf, í mismunandi keppnum á árunum 2018 til 2024. Mælaborðið var hannað til að sýna gögn á skýran og sjónrænan hátt með áherslu á samanburð keppenda á lykilatriðum eins og tíma, sætaröðun (Rank), fjölda keppna á ári og meðalhraða (pace) á kílómetra.

Hönnun mælaborðsins byggðist á eftirfarandi lykilatriðum:

- Samanburður á tíma og sætaröðun:** Með því að nota línurit er hægt að skoða hvernig Bergdís og Eyjólfur stóðu sig í keppnum þar sem þau kepptu bæði. Hér er notast við gögn eins og tíma þeirra í mínútum og sætaröðun þeirra í hverri keppni.
- Árangur á milli ára:** Mælaborðið sýnir breytingar á keppnistíma fyrir ákveðnar keppnir, eins og Íslandsmót í Criterium fyrir Bergdísi og Eyjólf, ásamt árangri Eyjólf í Fellahringnum.

3. **Fjöldi keppna á ári:** Með stöplariti er fjöldi keppna sem þau tóku þátt í á hverju ári sýndur, til að sjá hvort þátttaka þeirra hafi verið stöðug eða breytileg.
4. **Meðalhraði (pace) á kílómetra:** Reiknaður er meðalhraði þeirra (pace per km) í hjólræðiðakeppnum og borinn saman í línuriti. Þannig er hægt að greina hvort frammistaða þeirra hafi þróast til betri vegar eða breyst eftir keppnum.

Mælaborðið gerir notendum kleift að sjá gögnin á myndrænan hátt sem auðveldar greiningu og skilning á því hvernig þessi tveir hlauparar hafa staðið sig í keppnum. Notendur geta þannig notað mælaborðið til að fá heildaryfirsýn yfir frammistöðu, bera saman hlaupara, og skoða þróun þeirra yfir tíma.